

Better news — perhaps — on global warming

Evidence is emerging that carbon emissions may have started to fall, in advance of ratification of the Kyoto Protocol. If so, containing global warming may turn out to be less painful than has sometimes been supposed.

Recent surveys by the Worldwatch Institute, in Washington DC, and by the Energy Information Administration (EIA) of the US Department of Energy suggest that carbon dioxide emissions fell slightly last year — despite robust economic growth — both globally and in the United States (see page 494). While the data are insufficient to provide firm evidence of a clear trend, these findings at least give both sides of the global warming debate a rare opportunity to share the same piece of good news.

The EIA, an independent statistics unit in the energy department, found that carbon emissions held steady in the United States last year, and that industrial emissions dropped by 1.2 per cent, even as the economy grew at the exceptional rate of 4 per cent. The Worldwatch Institute, meanwhile, estimates that global emissions fell by 0.5 per cent in 1998, at a time when the world economy grew by 2.5 per cent.

The institute reports that China, whose emissions are widely recognized to be critical to any global effort to tackle the greenhouse gas problem, managed to cut them by almost 4 per cent, even as its economy boomed. All of this tentatively suggests that carbon emissions can be reduced without inflicting the economic damage that critics of the Kyoto Protocol say will be the inevitable consequence of rigid adherence to its targets for reductions.

It is prudent, of course, to note the statistics' potential limitations. The decline in US emissions may well be partially attributable to a

mild winter there last year, or to other extraneous factors. And it remains unclear whether China's reported drop in emissions is either accurate or sustainable.

At the same time, a number of developments that can be linked to the Kyoto process would seem likely to contribute to cuts in the growth of emissions. In particular, many large corporations have set their own targets, and have begun to implement them — even in advance of international ratification of the Kyoto agreement. Perhaps heightened awareness of global warming will, on its own, make a significant contribution to curtailing emissions growth — an important lesson in itself.

That would be just as well, since representatives of developed and developing countries are said to be making little progress towards settling their unresolved differences on the implementation of the Kyoto Protocol, and the Clinton administration clearly lacks the votes in the Senate to win its ratification in the United States.

More evidence that the connection between economic growth and emissions growth has been severed may be the best hope left for those who want more intensive action to forestall global warming. Such evidence would isolate hard-line opposition to emissions restrictions, cool the heated rhetoric of the global warming debate, and open the way for a relatively painless transition to a world of lower carbon emissions. □

Yes, minister, but ...

Britain's BSE inquiry is providing a valuable lesson in the relationship between science and public policy.

When the full history of Britain's handling of its crisis over bovine spongiform encephalopathy — BSE, or 'mad cow disease' — comes to be written, the transcript of the current inquiry being held under Justice Phillips will be a key text. And few passages of that transcript are likely to be as rich, particularly as an illustration of the highly complex relationship between science and politics, as last week's evidence by former and current members of the Spongiform Encephalopathy Advisory Committee (see page 490).

The committee was set up in 1990 on the recommendation of an earlier inquiry headed by Sir Richard Southwood, formerly professor of zoology at the University of Oxford. Its remit was a sensible and desirable one — or, at least, it appeared that way at the time: namely, to keep an eye on what was known and what was not known scientifically about the nature of BSE, how it was spreading among British beef herds, and its possible implications. In particular, these included the likelihood, previously described by the Southwood committee as 'remote', that BSE could spread from cattle to humans.

In practice, this remit appears to have been impossible to meet. Politicians and government officials, under intense pressure to provide statements that could be used as public justification of a particular course of action (or, just as frequently, inaction), perhaps inevitably sought to draw on the conclusions of a group of scientific

experts who could be held to be independent. Most notable was when John Gummer, the minister of agriculture at the time, invoked the name of David Tyrrell, then chair of the committee, when he stated that "What Tyrrell says, I will do".

It will be up to Phillips, whose inquiry is due to finish at the end of the year, to determine whether a crucial mark was overstepped in compromising the responsibility of a committee whose original brief was restricted to science alone. The present evidence indicates that this did indeed happen, although the full response from the civil servants and others involved is yet to be known. What is already clear, however, is that the inquiry has revealed the need for a much sharper definition of the role of science advisers in policy making.

The publication and propagation of guidelines on just this topic by Sir Robert May, the government's chief scientist, is an important step in the right direction. Even these, however, appear to argue for maintaining a sharp distinction between the scientific and the political. One of the key questions that the Phillips inquiry needs to answer is whether this distinction remains realistic in such sensitive (and scientifically uncertain) fields as BSE and its relationship to Creutzfeldt-Jakob disease, or whether new formulations can be imagined. Providing firm guidance on this alone will turn out to be a valuable contribution to modern politics. □



France to collaborate with UK on building new synchrotron

[LONDON] The French government announced on Monday (2 August) that it has agreed to play a major role in financing the construction and operation of a new third-generation synchrotron radiation facility in Britain.

The news has been welcomed by the British government and the Wellcome Trust, which last year agreed to provide £100 million (US\$165 million) towards the cost of the 3-GeV synchrotron machine, known as Diamond (see *Nature* 394, 209; 1998).

But the announcement has provoked an angry reaction in parts of the French research community, which had been lobbying for funds for its own planned 2.15-GeV third-generation synchrotron, Soleil.

This project had been seen by many as a key requirement for meeting France's declared aim of boosting research in the life sciences — particularly in exploiting genome sequence information. It was also regarded as the logical successor to ageing facilities at the Laboratoire de l'Utilisation du Rayonnement Électromagnétique (LURE) at Orsay, close to Paris.

The French announcement comes two years after the abandonment of earlier efforts to persuade British and French scientists and engineers to collaborate on the construction of a single synchrotron. At the time, it was argued that there was sufficient demand for access to synchrotron beams in both countries to justify two separate machines (see *Nature* 387, 539; 1997).

But it also follows a recent series of controversial statements by Claude Allègre, the minister of national education and research. Allègre has drawn the ire of many French researchers with his suggestion that, given the number of European countries now planning synchrotron facilities, there is no a strong case for building yet another machine in France (see *Nature* 395, 831; 1998).

In a statement issued in Paris on Monday, the ministry points out that several research facilities in France are used by researchers throughout Europe. It says the decision to participate in the British synchrotron is part of a broad strategy to promote the emergence of a European scientific community "competitive with the American or Japanese scientific community".

According to the ministry, France estimates that it will pay FF350 million (US\$56 million) towards the construction costs of Diamond over seven years, and FF60 million to FF80 million a year to its operating costs.

The statement adds that French scientists and engineers will be closely involved in the

design and construction of the synchrotron. France and Britain have, with the Wellcome Trust, agreed to set up joint technical committees that will draw up detailed specifications for the project and how it will operate, as well as detailed costings.

France will also set up a special laboratory on the Diamond site aimed, in particular, at welcoming French researchers. But the exact location of Diamond will be decided by Britain alone, and a decision is expected next month. The project is planned as a successor to synchrotron facilities at the Daresbury Laboratory in Cheshire.

The Wellcome Trust says it is "delighted" at the French involvement in Diamond. It says it is keen to ensure that the synchrotron needs of the biomedical research community are met. "The involvement of the French in the design of the facility will help to ensure that this is fulfilled."

A spokeswoman for the UK Department of Trade and Industry said that economies of scale from the collaboration would benefit both countries. "It is an excellent example of a most productive form of European collaboration," said the spokeswoman. "This will be a superb opportunity to build a world class facility, supporting leading-edge research in a wide range of disciplines — not only in the physical sciences but also and

notably in the biological sciences."

Reactions have been more muted — and frequently hostile — in France, where there has been a lengthy political stand-off between Allègre and several French regions, each keen to attract the facility.

Last week, for example, shortly before the government's decision was made public, leading scientists involved in the various bids to host Soleil issued a press statement warning that relying on foreign synchrotron facilities — and Diamond in particular — could be "disastrous for the long-term development of high-level research, fundamental and applied, in France".

The French government says that the continued operation of LURE is not under threat. But in June, François Wuilleumier, who headed the campaign to locate Soleil in the Paris region, issued a statement saying that France's participation in Diamond would only meet 15 to 20 per cent of the needs of French researchers.

Wuilleumier, a director of research at the Laboratory of Atomic and Ionic Spectroscopy of the National Centre for Scientific Research, said that not building Soleil would be a "fatal blow". The harmful consequences would end a scientific adventure "in a field in which France has been among the pioneers for the past 30 years". **David Dickson & Natasha Loder**

Wallace rescued from a grave injustice

[LONDON] A group has been set up to raise funds to restore the neglected grave (right) of Alfred Russel Wallace, who discovered evolution by natural selection independently of Charles Darwin.

Darwin was buried in Westminster Abbey in London, alongside royalty and figures such as Isaac Newton, and is remembered as one of the world's greatest scientists. Wallace, in contrast, has been almost forgotten by history, although some believe the question of the priority of the discovery remains open.

Wallace's grave has been neglected for many years in a cemetery in Broadstone, Dorset. A striking feature of the grave is a seven-foot fossilized conifer trunk, whose significance is unknown, but which is now in need of attention.

George Beccaloni, a researcher at the Natural History Museum in London, located the grave after a search last year. "I was quite horrified by its poor condition," he says. Along with Wallace's grandson Richard Russel Wallace, Beccaloni has launched a



memorial fund to restore and protect the monument.

They hope to gain legal protection for the grave and to extend its lease and transfer it to the Linnean Society. They also want to secure a plaque for Wallace's house, "The Dell", in Essex, where he lived from 1871 to 1876 and wrote *The Geographical Distribution of Animals*. **Natasha Loder**

Congress talks tough on funds for NASA and basic research

[WASHINGTON] Science lobbyists in the United States face a tough battle this summer to restore research funding at the space agency NASA and at the National Science Foundation (NSF). The House of Representatives' Appropriations Committee has proposed a bill that would slash NASA's budget by almost \$1 billion and freeze spending at the NSF.

Efforts will be made to restore some of the NASA spending this week, when the Republican-controlled House tries to pass the broad budget bill that covers Veterans' Affairs and Housing and Urban Development (VA-HUD) and independent agencies, which includes funding for the NSF and NASA.

A spokesman for Jim Sensenbrenner (Republican, Wisconsin), chairman of the House Science Committee, said he was "trying to ensure that some of the funds were restored" before the House votes on the bill.

The version of the bill passed by the Appropriations Committee last week proposed huge cuts in research at NASA. The bill would reduce the space agency's total budget from \$13.7 billion to \$12.7 billion, and cut space science by \$240 million and Earth science by \$285 million.

The committee voted to restore \$400 million worth of cuts — including the cancellation of the Space Infrared Telescope Facility project — which had earlier been endorsed by its VA-HUD subcommittee.

The bill would also hold the NSF's budget at its 1999 level of \$3.65 billion, denying a request from the Clinton administration for a 6 per cent increase. The version would grant only \$35 million of the \$146 million that the administration requested for a new information technology research initiative.

Agency heads moved quickly to attack the proposals. "Not only is this cut devastating to NASA's programmes, it is a knife in the heart of employee morale," said NASA administrator Dan Goldin. NSF director Rita Colwell said in a statement: "We're ready to do twenty-first century science and engineering, but we can't do it on a twentieth-century budget."

But the cuts proposed in the House VA-HUD bill are a long way from being implemented. They reflect the low budget allocation granted to the bill to keep spending under the tight budget caps agreed by the Congress and the administration in 1997.

The cuts proposed for NASA's Earth sciences programme will be bitterly opposed by influential senators such as Barbara Mikulski (Democrat, Maryland), whose state is home to the Goddard Space Center. The White House also said it would veto the bill on various grounds, including its funding levels for NASA and the NSF.

Colin Macilwain

BSE advisers admit giving up a purely scientific role

[LONDON] Scientists advising the British government on the outbreak of the bovine spongiform encephalopathy (BSE) epidemic in the late 1980s claim that they came under pressure from officials to endorse a statement on the safety of beef. They also admit they were uncomfortable about their relationship with civil servants and ministers.

The admissions emerged last week during the questioning of members of the Spongiform Encephalopathy Advisory Committee (SEAC), at the public inquiry headed by Lord Phillips.

SEAC has handled advice to the government on the epidemic since 1990. It was supposed to stand back from immediate issues and focus on underlying science, but this goal appears to have been compromised only two-and-a-half weeks after it first met.

SEAC members were contacted by the Chief Medical Officer (CMO) on 16 May 1990 and asked to approve a statement he was issuing the following day saying that beef was safe. David Tyrrell, who headed SEAC at the time, said its members had felt uncomfortable and anxious about having to approve a statement at such short notice.

At the inquiry, SEAC members were questioned on the committee's close working relationship with the government. They said that a shortage of time to deal with pressing problems and a shortage of technical expertise had left SEAC working closely with the Ministry of Agriculture, Fisheries and Food (MAFF).

Phillips and others conducting the questioning also expressed interest in the way that the government's CMO was kept closely informed of SEAC's deliberations by Hilary

Pickles, a member of the committee's secretariat provided by the Department of Health.

Pickles drafted material for SEAC, as well as for an earlier scientific committee advising the government on the BSE outbreak (see *Nature* 400, 389; 1999). Pickles put together material for the CMO's appearance in front of a government committee — and this later formed the basis of SEAC advice to the CMO.

It emerged that the draft SEAC report was also circulated to MAFF officials for comment. Asked whether he saw a possible conflict in the fact that departments expecting advice were contributing to it, Tyrrell said, "We had already sold the pass, having said we are going to be involved in doing things to help a CMO".

He added: "We had given up the idea of trying to stand back and do nothing else but evaluate science at a distance and impartially."

Tyrrell later rephrased his words, saying the committee members would have picked up changes to the document that they disagreed with. But he admitted to being worried about dropping from their advice the phrase: "No scientist would say there was no risk of eating beef". MAFF officials felt that certain statements were inflammatory, said Tyrrell, and were worried about public reaction.

Asked if the draft paper should have been subjected to this kind of process, Tyrrell said that with hindsight and in an ideal world this should not have happened. He would defend the situation at the time, however, "by saying it was so fraught, almost that we had to do something, or at least it was seen by those in the ministry that we had to do something and get it out soon".

Natasha Loder

UN call for action to clear up space junk

[MUNICH] Member states of the United Nations (UN) have been asked to take steps to increase awareness of the potential impact of space activities on science, and on the economic development of rich and poor nations. Such steps should include exploring the legal aspects of space debris, and protecting some regions of the Earth from radio emissions.

The recommendations are included in the Vienna Declaration on Space and Human Development and a related action plan that were approved last week at the end of the third UN Conference on the Exploration and Peaceful Uses of Outer Space — UNISPACE III — held in Vienna, Austria.

Those attending the conference also urged the creation of a voluntary UN fund

to translate the recommendations into action. The two-week meeting was the first UN space conference since 1982. In its technical forums, astronomers expressed serious concern that light pollution, radio interference and space debris are endangering their research.

The delegates recommended holding international meetings of researchers on near-Earth objects, organized by the UN Officer for Outer Space Affairs. They also called for the creation of an international space authority to facilitate cooperation on space debris. This organization should also coordinate the detection of landmines from space, and develop legal frameworks.

The declaration stresses that Third World countries must not be excluded from the benefits of space science. Eva von Schaper

Asia backs E-Biomed 'with peer review'

[TOKYO] The Asia-Pacific International Molecular Biology Network (IMBN) last week endorsed in principle the proposal for a global repository of biomedical literature — known as 'E-Biomed' — put forward by the US National Institutes of Health (NIH).

But the organization, which brings together life scientists in the Asia-Pacific region, followed the lead of the European Molecular Biology Organization (EMBO) in reacting cautiously to the suggestion that the repository should contain unrefereed submissions. It stressed that the interpretation of biological research results requires critical review from a third party.

At their second international conference, held in Singapore, IMBN members backed a motion by Frank Gannon, executive director of EMBO, on which IMBN is modelled, which last month indicated its intention to provide European input into E-Biomed (see *Nature* 400, 97; 1999).

Gannon has indicated that E-Biomed should have an international governing board represented by the NIH, EMBO and an Asian entity. IMBN, which represents 14 countries in the region, including Australia, China, Indonesia, Korea, Japan and Singapore, was considered the most appropriate.

"IMBN will support the concept of E-Biomed and the approach EMBO is taking

with the planned initiative," says Gurinder Shahi, executive director of the IMBN. But he admits that "the devil is always in the detail", pointing out, for example, that one of EMBO's main concerns over E-Biomed is the risk of losing income through its journals.

"This is a slight worry for us too, as we are planning to launch our own journal, including an electronic-only version, between 2000 and 2001," he says. "But the priority is to recognize the importance of Asian participation in E-Biomed. The most important thing is to ensure IMBN's inclusion in E-Biomed at an early stage, so that we could provide a direction in the initiative."

Many Asian life scientists say they welcome the concept of the planned repository and would benefit from its simple, fast and barrier-free system. They say it would be particularly attractive to scientists in developing countries, as it would facilitate timely access to materials and information that are currently too expensive to subscribe to or are significantly delayed in their delivery.

"We also hope that E-Biomed would help to break the strong dependence on 'high impact factor' journals in Asian countries for assessing research performance," says Chris Tan, director of the Institute of Molecular and Cell Biology in Singapore.

Tan predicts that the repository would

narrow the gap between so-called 'brand name' journals, such as *Nature* and *Science*, and other journals that cover more specialized areas of science. He says that a system such as E-Biomed would help to liberate knowledge currently confined to specialized journals and encourage its distribution to Asian researchers.

But he is also concerned about the absence of peer review in the E-Biomed proposal. "The downside of E-Biomed is the risk of attracting too much garbage if the peer-review system is not at work. Although [the] system suffers from slow review times, I feel that it should be retained," says Tan.

"Submissions of papers from the Asia-Pacific region, particularly from developing countries, would be greatly encouraged by E-Biomed, which would provide a more accessible platform for publication," says Jerry Wang, professor of biochemistry at the Hong Kong University of Science and Technology. But he also agrees that some form of quality control should operate.

IMBN may push for a wider range of life-science literature to be included in the repository, as it now recognizes new areas of interest such as agricultural and environmental sciences. This may already be under way, as there is talk of changing the name from 'E-Biomed' to 'E-Biosci'.

Asako Saegusa

Varmus speculates on a possible reorganization of the NIH

[BAR HARBOR, MAINE] Harold Varmus, the director of the US National Institutes of Health (NIH), said last week that it was time to discuss how the huge biomedical agency might be remodelled to increase its effectiveness — for example, by consolidating it into a smaller number of independent institutes and giving an expanded director's office more power.

Speaking at a symposium on the future of genetics organized by the Jackson Laboratory, of Bar Harbor, Maine, and the Johns Hopkins University, Varmus said that the NIH should aim to "organize the science in a rational way". Rather than continuing to subdivide the \$15.6 billion agency — which since its founding in the 1930s has grown into more than two dozen institutes, centres and divisions — he said that more might be achieved by having fewer centres of power.

Under one proposal, put forward, he said, purely for discussion purposes, there might be just five institutes — based on research into cancer, neuroscience, general medical sciences, human development, and microbial and environmental sciences.

A proposed brain institute would encompass the activities of the National Institute of Mental Health, the National



Varmus (top right) addresses a recent hearing in Congress surrounded by young diabetes sufferers. Should research into the disease remain the responsibility of a separate institute?

Institute of Neurological Disorders and Stroke, the National Institute on Alcohol Abuse and Alcoholism, and the National Institute on Drug Abuse. A revised National Institute of General Medical Sciences would include the work of the National Institute of Diabetes and Digestive and Kidney Diseases and that of the National Institute of Arthritis

and Musculoskeletal and Skin Diseases.

Under this scheme, a transformed Office of the Director, perhaps called 'NIH Central', would oversee the institutes. This would direct policy, oversee the Library of Medicine and the Center for Information Technology, and support training, peer review and activities not handled by the institutes.

Varmus said that the central office should have broader responsibilities and a budget enabling it to support special initiatives, such as developing technologies relevant to research across the institutes. It should also have the flexibility to redirect funding on a project's completion.

The comments came amid rumours that Varmus, who has directed the NIH since 1993, is being considered to head the Memorial Sloan-Kettering Cancer Center in New York. The *New York Observer* last week quoted the director of another New York cancer centre as saying that the hiring of Varmus is "close to being done".

Varmus will not confirm or deny reports that he is leaving the NIH. Anne Thomas, his spokeswoman, says that Varmus "has had conversations with Memorial Sloan-Kettering, and with other institutions. He has not made a final arrangement with any."

Carina Dennis

NSF urged to raise environmental efforts ...

[WASHINGTON] The US National Science Foundation (NSF) should increase its annual spending on environmental research from \$600 million this year to \$1.6 billion in five years' time. So concludes an interim report that was approved last week by the foundation's governing body, the National Science Board (NSB).

An NSB panel concluded that the foundation's support for basic research into environmental questions "represents only about one-third of the resources necessary" in view of "the overwhelming importance and exciting opportunities for progress in the environmental arena". The panel was chaired by Jane Lubchenco, a marine biologist at Oregon State University.

If implemented, the call for extra money would radically reorientate the NSF's \$2.5 billion university research portfolio. The report is the NSB's main response to an instruction in 1997 from the Congress that the foundation should put together a plan to establish and run a National Institute for the Environment (NIE).

But the report rejects the idea of a separate NIE or an environmental directorate within the foundation. It calls instead for a "high-visibility, NSF-wide organizational focal



Lubchenco: 'no need for a new directorate'.

"It's time for a major increased effort in environmental science at NSF," says Lubchenco. "An expanded effort is necessary, possible, and will be very exciting." Lubchenco adds that the foundation "will need a new focus on interdisciplinary research".

Eamon Kelly, an economist at Tulane University in Louisiana and NSB's chair, says the extra money is "a very conservative estimate" of what is needed. He adds that congressmen of various political perspectives want to see

point" for environmental research. "A new directorate is not necessary," says Lubchenco.

This finding disappoints advocates of an NIE, who believe that the NSF will be unable to support interdisciplinary environmental work unless it augments its structure of discipline-specific directorates and divisions.

more environmental research at the NSF.

The report recommends that the NSF should pursue more research into methods that support environmental assessment and education, and should create environmental research centres at US universities. It should also research better information networks for environmental scientists and support research into environmental technologies, says the report.

But NSF director Rita Colwell stresses that the proposals would stick to basic, rather than applied, research. "The NSF is uniquely placed to lead in basic environmental research," she says.

Research into the genomes of microbes, for example, falls within NSF's remit. But the gathering and processing of environmental data do not. "We wouldn't be monitoring acid rain," says Colwell. "Our sister agencies, such as the Environmental Protection Agency (EPA), do that."

But campaigners for the NIE would like the NSF to do exactly that. They include industry groups that do not trust the EPA's research, and liberal academics who believe the EPA and other agencies do not have enough resources for research.

The Campaign for a National Institute for the Environment (CNIE), which persuaded Congress to ask the NSF to establish an NIE, has drawn support from the US Chamber of Commerce and many leading scientists. Until she became NSF director last year, Colwell was on the campaign's board of directors.

"The NSB has produced a landmark document," says Richard Benedick, a retired diplomat who is president of the CNIE. "But the real proof of its value will come from how the NSF implements its recommendations." Benedick says that the NSF "was being urged to support both basic and applied research", and it is "regrettable" that it has not backed a new directorate for environmental science.

Ron Pulliam, a professor of ecology at the University of Georgia and former head of the disbanded National Biological Survey at the Department of the Interior, says the report is "very exciting and very promising".

But Pulliam, who is on the CNIE board, suggests that the NSF must look beyond basic research and do some of the environmental assessment work that would support the missions of the rest of the government. "There's a feeling that this is not being done very well, and Congress has been in no mood to give the EPA or the Department of the Interior the money to do it."

Last week's report makes clear that the NSF does not want to do it either. The foundation is proud of its dedication to basic research and fearful of becoming embroiled in controversial policy decisions about environmental regulation.

Colin Macilwain

... as panel backs the return of the 'green GDP'

[WASHINGTON] A panel of the National Research Council (NRC), part of the US National Academies, has recommended that the US government resume attempts to measure the nation's green gross domestic product, known as 'green GDP'.

Such a measure includes environmental resources in estimates of economic productivity. But the US Congress is unlikely to authorize any work in the near future, having stopped it five years ago.

The Commerce Department's Bureau of Economic Analysis (BEA) began developing environmental accounting methods in the early 1990s, and in 1994 published its first analysis, of subsoil mineral resources.

That same year, two congressmen from coal-producing states, Harold Rogers (Republican, Kentucky) and Alan Mollohan

(Democrat, West Virginia), led a move to stop work on green GDP until the NRC could assess the methods and objectivity of the department's analysis.

The panel, led by Yale economist William Nordhaus, concluded in its report, released last month, that the BEA's work was sound, and that "developing a set of comprehensive non-market economic accounts is a high priority for the nation".

Rather than fold environmental factors immediately into the core, market-based GDP account, however, it suggests creating a non-market "satellite account". The panel also calls for a "concerted federal effort" to measure changes in the quantity and quality of environmental assets – data that are sorely lacking.

It would cost about \$1.5 million annually to resume research on green GDP, according to the BEA. But

bureau director Steven Landefeld says that, even with the NRC's endorsement, "I'm not sure anything will happen quickly on this". His budget for conventional economic analyses has been lean in recent years, and he estimates that only one person could be assigned to green GDP research part-time if it started up again.

Landefeld says the United States once led in this type of statistical analysis, but has been out of the debate for several years. With the NRC report complete, he says there may be another opportunity for US leadership.

Nordhaus specializes in modelling the economic effects of climate change. The NRC staff behind the green GDP study offered for him to brief the commerce appropriations subcommittee that commissioned the report. So far, though, no one has taken up the offer. **Tony Reichhardt**

Charity cools on stem cells after boycott by Catholics

[WASHINGTON] There was a new twist last week to the debate over whether the US government should support research that uses human embryonic stem cells. It became known that the American Cancer Society (ACS) has withdrawn from a coalition that is lobbying Congress to support such research.

The society, which raises \$500 million a year, earlier this month asked the Patients' Coalition for Urgent Research — known as 'Patients' CURE' — to remove its name from the list of 31 research- and patient-advocacy groups that make up the coalition.

The ACS had come under pressure from officials of the Catholic church not to participate, while some 100 lay Catholics had withdrawn from a breast-cancer fund-raising event in the suburbs of Washington and revoked promised contributions.

The current controversy was triggered by the first report last November that human embryonic stem cells had been isolated (see *Nature* 396, 104; 1998). Although there is a congressional ban on federal funding for research in which human embryos are destroyed or discarded, the Department of Health and Human Services ruled that federal funding of research on stem cells is permissible provided their derivation is privately financed and is limited to the use of spare embryos from fertility clinics (see *Nature* 397, 185–186; 1999).

This ruling has drawn protests from anti-abortionists on the grounds that it is a violation of the congressional ban to allow the government to fund research with cells that depend for their production on the destruction of embryos. Patient-advocacy groups, in contrast, support the research, which could lead to cell and tissue therapies for a range of diseases, including juvenile diabetes and Parkinson's disease.

The Patients' CURE coalition was launched in May to lobby Congress to back government funding of embryonic stem-cell research through the National Institutes of Health (NIH). The ACS's withdrawal — first reported in *The New York Times* — suggests that emotions are running so high that even solid research-backing organizations are feeling the heat from constituents who oppose abortion. But Greg Donaldson, a spokesman for ACS, denies that the society has reversed its position.

Donaldson says that ACS is only now formulating a position on stem-cell research, and that therefore "it is impossible for us to have reversed our position". He says that, although press material passed out at the coalition's launch included ACS among its

members, the society never formally joined Patients' CURE, but simply "stood up with" it at its launch when Congress seemed about to prohibit the federal funding of stem-cell research.

The withdrawal from the coalition followed a letter on 18 May from William Cardinal Keeler of Baltimore, who heads the Secretariat for Pro-Life Activities at the National Conference of Catholic Bishops. The letter urged the ACS not to participate in the group.

Also in May, around 100 people in the Maryland suburbs of Washington withdrew from ACS's breast-cancer fund-raising event Relay for Life shortly before its launch. The group, most of them Catholic, had learned of ACS's membership of Patients' CURE, and said they could not support the society as long as it backed stem-cell research. Pledges worth several thousands of dollars were returned.

One member of the group, Tim Flynn, a private investment manager in Silver Spring, Maryland, whose wife lost her mother and 39-year-old sister to breast cancer, says that "my family and my wife's family and a large group of friends very desperately want to see progress made in cancer research and cures for a number of illnesses. [But it is] inconsistent that you would destroy life for any purpose, certainly scientific experimentation. You can't use the 'ends justify the means' rationale."

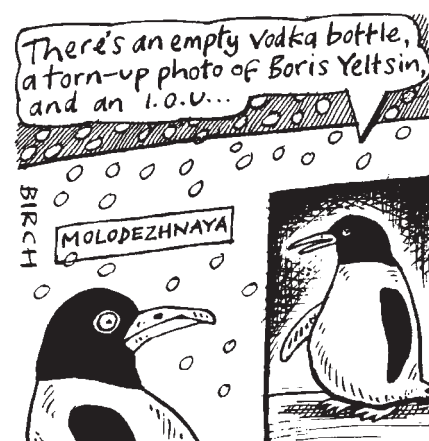
Daniel Perry, a spokesman for Patients' CURE, says that other members of the coalition have been under pressure to withdraw but have resisted. "Some groups apparently can tolerate the heat a little better than others," he says. "We're all getting hate mail."

He says that the fact that two groups have recently joined the coalition is evidence that its opponents' strategy is failing. The new groups are Hadassah, the Jewish women's group, and Prevent Blindness America.

Also last week, a senior Republican senator and a Catholic Republican congressman held a press conference with Patients' CURE to endorse federal funding of stem-cell research. Senator Strom Thurmond (Republican, South Carolina), whose daughter has juvenile diabetes, urged Congress to allow federal funding of the research. The NIH "must be actively involved in stem-cell research", he said.

And Brian Bilbray (Republican, California), who describes himself as pro-choice, declared that it was "outrageous" for Congress to tell the parents of spare embryos stored at infertility clinics that they could not "save a life with an embryo that's not going to be born".

Meredith Wadman



Russian Antarctic base falls prey to budget cuts

[MOSCOW] Russia's main Antarctic scientific base, Molodezhnaya, has been closed owing to lack of funds. The fate of the base, which has been shut "until better days", was sealed two years ago, when the Russian cabinet adopted a strategy of adapting Antarctic research to meet budgetary pressures.

Molodezhnaya, founded in 1963, was a heavy burden on the limited budget, as it has the best-developed infrastructure of Russia's Antarctic bases, including permanent living quarters for scientists, whose upkeep required large amounts of money.

Furthermore, as Valery Martysenko, head of the the Russian Hydro-meteorological Service (RHMS), points out, the finance minister has delivered only one third of the 72.5 million rubles promised for the station.

The closure of the base means that only four of the Soviet Union's original eight stations in Antarctica are left: Mirny, Progress (which now becomes Russia's main research centre), Vostok and Novolazarevskaya. Each is supposed to continue research and fulfil Russia's commitment to the international programmes in which it participates.

But there could be difficulties in closing the Molodezhnaya base because, under the terms of the Antarctic Treaty, no waste can be left behind. "For this, we need special equipment, containers and many other items that we are unable to buy," says Sergei Khodkin, RHMS deputy head.

Ships must leave Vladivostok no later than October to complete the clean-up before the Antarctic ice becomes too thick. But the cabinet is not scheduled to consider the problem until the first week of September.

Russian ecologists warn that international organizations could employ a private contractor to clean up the Molodezhnaya site, and then send Russia the bill, perhaps with an additional fine. If so, the attempts to save money could end up costing more than keeping the base open.

Carl Levitin

US plans giant effort on protein structure

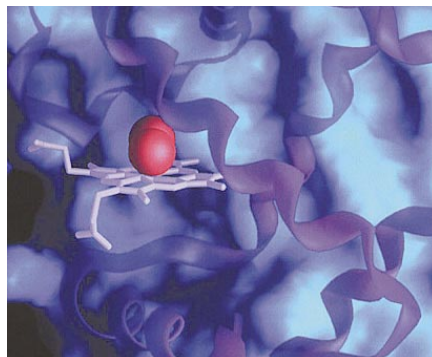
[WASHINGTON] US scientists have set out to compile a comprehensive catalogue of protein structures. The idea takes its cue from the Human Genome Project's bid to decipher the entire human genetic code.

The undertaking, dubbed structural genomics, has been given a boost by the announcement that the National Institute of General Medical Sciences (NIGMS) is to launch a Protein Structure Initiative. This will spend up to \$3 million a year at each of three to six pilot research centres.

These centres will explore the best ways to determine protein structure quickly, accurately and at high volume. The ultimate goal is to deduce at least 10,000 protein structures within the next five years, at the lowest possible cost. At present it costs about \$200,000 to determine a structure. It is hoped that technical improvements and economies of scale could bring this down to \$20,000.

It would be impossible to catalogue every protein. But, by choosing those that do not show sequence similarity to proteins of known structure, structural biologists hope to find novel forms. The aim is to comprehensively describe a finite set of what are thought to be several thousand basic protein shapes, or folds, and variations on them. Currently, only a few hundred are known.

Once this task is complete, it should be possible to assign any protein to a shape family simply by knowing its genetic sequence. Once the detailed structure of one member of a family is known, computer modelling will make it possible to generate a reasonably good structure for any other member.



Shape of things to come? A graphic of the protein myoglobin, carrying oxygen (red).

It is hoped that the project will be a starting point for understanding how virtually any protein works. It could yield information on protein evolution, the basic physics of the relationship between protein sequence and structure, and structure-based drug design and discovery.

This would have a tremendous impact on pharmacology. At the moment, only about one per cent of all protein families are targeted by drugs. But drug development is greatly helped by a knowledge of protein structure. The selection of targets for drugs can become more rational, and drug candidates can be tailored for maximum effect on their targets.

The NIGMS-funded effort was hatched after meetings over the past year between structural biologists and officials at the National Institutes of Health. They have sought to capitalize on two events. One is the explosion of sequence data being generated

by the Human Genome Project. The second is the improvement in technologies for studying protein structures. For instance, highly focused, bright X-ray beams, called undulator X-ray sources, at synchrotrons, allow data to be collected from smaller crystals, and in minutes rather than days.

It is hoped that the pilot centres will become expert in deducing protein structures. This would involve selecting and producing target proteins, crystallizing them, determining their structures and analysing the data generated. The information would be stored in a common database.

After three years of developing and refining technologies and testing strategies for producing speedy, high-volume results, the less productive centres might be culled. Those remaining would go into high-volume structure determination, producing most of the structures in the following few years.

Ultimately, large, integrated centres are planned, and NIGMS seeks collaborations with Europe and Japan. Marvin Cassman, NIGMS director, says the focus is on scale and speed. The key is "rapid structure determination at high resolution and high throughput," he says. "This is intended ultimately to be a kind of assembly-line process."

The plan is winning plaudits from many researchers. They say that a comprehensive compendium of protein structures will be a boon to biology, and that a large-scale, systematic approach is the way forward.

"It will be great," says Andrej Sali, an assistant professor in the molecular biophysics laboratories at Rockefeller University in New York. "It will have an impact at least as large as the optimists are predicting, just as the Human Genome Project did."

But critics echo early objections to the genome project, complaining that the plan will channel valuable resources to an exercise of dubious value. "Trying to figure out function from structure is one of the most difficult enterprises in molecular biology," wrote Thomas Steitz, a Howard Hughes Medical Institute investigator at Yale University, in a letter to Cassman.

Steitz is a professor of molecular biophysics and biochemistry. He says that structural biologists ought to be left to do what they are already doing — determining the structures of proteins known to be biologically important. "Most of us could come up with a long list of far more important and interesting projects," he wrote to Cassman.

But others are more impressed. It "could give databases that could be extraordinarily valuable and time saving in research efforts," says Wayne Hendrickson, a Howard Hughes investigator in the Department of Biochemistry and Molecular Biophysics at Columbia University, New York.

Meredith Wadman

Emissions fall despite economic growth

[WASHINGTON] Global carbon emissions declined last year while economic productivity increased, providing evidence, according to the Washington-based Worldwatch Institute, that cutting greenhouse gas emissions may not necessarily have a financial cost.

The institute, extrapolating from industry reports of fossil-fuel consumption, estimates that emissions fell by 0.5 per cent to 6.32 billion tons in 1998. Meanwhile, the global economy grew by 2.5 per cent.

It says that the numbers are a "sign that it may be less difficult to slow global warming under the Kyoto Protocol than has been assumed by some industry groups". Industry groups, in contrast, say the data vindicate their preference for voluntary reductions.

The pattern of falling emissions and economic gain was strongly evident in China, the world's second-largest emitter of carbon. The Chinese economy grew by

7.2 per cent last year, but its emissions fell by 3.7 per cent. The reasons for the decline are unclear, although the government has reduced subsidies for coal and has banned coal burning in Beijing homes.

Similarly, Poland's emissions fell by 9.7 per cent while its economy grew by 6 per cent. Emissions had already declined in central Europe in the early 1990s, but this was owing to economic collapse, not growth.

Although no one can be sure that the 1998 figures are a trend, several factors may be involved, says Worldwatch, notably energy efficiency, fewer fossil-fuel subsidies by governments, and the transition to a cleaner information economy.

The Worldwatch analysis follows the release of figures by the US Department of Energy showing that emissions in the United States rose by 0.4 per cent in 1998 while the economy grew by 3.9 per cent. The United States is still producing greenhouse gases at 10 per cent above 1990 levels. **Tony Reichhardt**

Africa seeks laws on GM food exports...

[LONDON] African countries are to be asked to introduce legislation that would make it illegal for a country to export genetically modified (GM) food without first seeking permission from the importing country.

The move comes in response to the collapse of talks on an international biosafety protocol earlier this year (see *Nature* 398, 6; 1999). African countries had wanted to insert a clause to this effect in the protocol, but this suggestion was strongly opposed by a consortium of grain-exporting countries led by the United States, and known as the Miami group.

These countries argued that a product that has passed safety tests in its country of origin should not need permission — or additional risk assessments — from an importing country.

The draft Africa legislation, which has been drawn up by a small group of countries led by Ethiopia, says that any person or organization intending to export GM food or use GM organisms in a laboratory or for field trials must carry out an evaluation of the risks to the environment, biological diversity and human health. Such an evaluation will also include socio-economic risks, such as the impact on jobs.

The draft says that permission will only be granted if the GM product does not pose risks of “adverse socio-economic impacts”, and if it “accords with the ethical values and concerns of communities and does not undermine traditional knowledge and technologies”. Under the draft, an exporter will also be liable to pay compensation for any harm caused by the GM material.

“No approval shall be given unless there is firm and sufficient evidence that the GM organism or the product of a GM organism poses no risks to the environment, biological diversity or health. Where there is reason to suspect threats of serious damage, lack of scientific evidence should not be used as a basis for not taking preventive measures,” says the draft legislation.

The final draft is expected to be completed next year when individual governments will be lobbied to adopt the proposals. “A few African countries have now decided to take the initiative and introduce national biosafety legislation in as many countries as possible,” says a source close to the Ethiopian government. “We don’t necessarily need to wait for an international protocol.”

Introducing Africa-wide legislation, however, is fraught with difficulties. Last year, for example, heads of state of the Organization of African Unity (OAU) agreed to adopt a continent-wide law banning patents on natural products until the World Trade Organization modifies its rules on patenting in line with the United Nations Biodiversity

Convention. But so far there has been no follow-up in any country.

One reason is that the OAU has little practical influence in the continent. Another is that trade and foreign ministries in Africa will not support any proposed law likely to antagonize France — which maintains considerable influence in the continent — or the United States. Neither ministry is likely to agree to a set of measures that are a direct challenge to the international trading system.

Countries belonging to French-speaking Africa have already agreed in principle to

accede to the UPOV convention (see *Nature* 398, 99; 1999), which grants plant breeders intellectual-property rights over the commercialization of products such as seed. English-speaking countries in Africa are expected to follow suit.

At a meeting of OAU heads of state last month, Salim Ahmad Salim, the secretary-general of the OAU and a former prime minister of Tanzania, appealed for countries to adopt legislation making it compulsory for companies that have developed medicines from natural products to share the benefits with local communities.

Ehsan Masood

... as Brazilian scientists protest at GM ban

[PÔRTO ALEGRE, BRAZIL] Scientists attending a meeting of the Brazilian Society for the Progress of Science (SBPC) in Pôrto Alegre, the capital of the state Rio Grande do Sul, have condemned the state government’s proposal to ban all genetically modified (GM) crops.

Francisco Mauro Salzano, a prominent geneticist at the Federal University of Rio Grande do Sul, described the state government’s attitude as “medieval”, and compared it to Stalin’s support of the agronomist Trofim Lysenko, whose ideas about genetics put back Soviet agriculture for many decades.

But at the opening ceremony of the SBPC meeting last month, the state’s left-wing governor, Olívio Dutra, said that in his state “we do not want Stalin’s dictatorship, the same way that we don’t want the dictatorship of Monsanto”. The move to ban GM crops from the state has yet to be passed by the state legislature.

Monsanto has recently applied to the federal government to import and commercialize soy seeds that have been genetically modified for resistance to the company’s herbicides. Brazil’s biosafety legislation regulates both the release of GM organisms into the environment and the sale of GM foods through the



Doctored label: black beans are used in a GM health scare.

National Technical Commission for Biosafety (CTNBio).

But even though CTNBio has given the green light to Monsanto, environmental and consumer groups are mounting a legal challenge to the import of seeds. Critics of these groups suspect a political agenda, however. A leaflet distributed by the state government shows a bag of black beans — a staple part of the Brazilian diet — with the label, in English, “original American seed” beneath an American eagle, and another, in Portuguese, that reads “for sale only with medical prescription”.

The environmental group Greenpeace has sought to turn Rio Grande do Sul into a “transgenics free state”. And a Rio Grande organization of landless peasants has vowed to burn any soy fields using GM seeds.

But Salzano’s criticism of the state policy has the support of many Rio Grande researchers. “The debate is being carried in a doctrinaire,

even fanatical way,” says José Francisco Valls of the agriculture research institute Embrapa, which brings together the agricultural research units of the Brazilian Ministry of Agriculture.

“There is paranoia here regarding transgenics,” says Maria Irene Baggio, also from Embrapa. She has a leaflet distributed by a local syndicate describing GM seeds as “the seeds of death”. “Aggression towards scientists is constant in these debates,” she says.

The only dissenting voice came from SBPC itself. Aware that the society has potential supporters of the ban among its members, it preferred to issue a more moderate statement urging caution and requesting a five-year moratorium in the use of GM seeds. During this period, more tests could be conducted in Brazil, according to SBPC’s new president, Glaci Zancan, a biochemist from the Federal University of Paraná.

The state government of Rio Grande do Sul hopes to capitalize on European concern about GM foods by offering to export to Europe soy production labelled as GM-free. One leaflet, in English, says the government agrees that science should be “under public control to benefit life, not under private control to [benefit] profit”. **Ricardo Bonalume Neto**

European Patent Office gives the go-ahead on transgenic organisms

[MUNICH] The European Patent Office (EPO) has adopted new rules on the patenting of biotechnological inventions, in line with a European Union directive approved last year (*Nature* 393, 200; 1998). The rules — which explicitly allow the patenting of transgenic plants and animals, as well as human genes of known function — will come into effect next month.

The move will allow the patent office to start work on a backlog of up to 100 applications for patents on animals and plants. Their processing has been suspended since 1995, when a board of appeal interpreted the EPO's rules on biotechnological inventions to preclude the patenting of animals and plants (see *Nature* 374, 8; 1995).

But this may not be the end of the issue. The EPO's highest legal authority, the Enlarged Board of Appeals, is currently considering how the principles relate to a rejected application by Novartis for a herbicide-resistant plant. It could overturn the rules if it finds them inconsistent with the 1973 European Patent Convention, the EPO's basic legal guidelines. The EPO would

then have to draw together its 18 signatory states to redraft the convention and ratify it through their national governments.

Stamp sales raise \$8m to help get cancer licked

[SAN DIEGO] About \$7.7 million has been raised for breast cancer research in the United States from the issue a year ago of a special postage stamp. The National Institutes of Health is to receive 70 per cent of the funds, with the rest going to the US defence department's medical research programme. California led the way in purchases of the stamps.

Call for crackdown on Israeli software pirates

[JERUSALEM] Israel must take stronger action against software and CD piracy, Israel's new industry and trade minister, Ron Cohen, told his cabinet colleagues last week. Cohen said that the country is already on the US blacklist for intellectual-property violations.

If action is not taken, he said, this will lead to US sanctions by next December, costing tens of millions of dollars. The minister called for harsher penalties for offenders, and for extra funds for the intellectual-property police unit, which has only 12 police officers.

\$20m grants create five US science centres

[WASHINGTON] The US National Science Board has approved five new, interdisciplinary science and technology centres at US universities — the first such centres to be created since 1991.

The centres will be established at Cornell University, New York state, for nanobiotechnology; the University of California at Santa Cruz, for adaptive optics; Emory University, Atlanta, for behavioural neuroscience; University of North Carolina, for environmentally-sound chemical processes; and the University of Arizona, for water resources.

Each centre will receive \$20 million in support from the National Science Foundation over five years. The centres will also involve several partner institutions. They will develop educational opportunities at high schools, as well as at universities, while pursuing interdisciplinary research.

Internet takes eclipse watchers to Turkey

[SAN DIEGO] The Internet will allow professional and amateur scientists worldwide to observe the Solar eclipse over Europe and Asia on 11 August, when a team

of San Francisco scientists will broadcast the event live from a mountain-top in Turkey. The full Solar eclipse will occur at 13:30 local time in Turkey and can be seen on www.exploratorium.edu/eclipse.

The US space agency NASA's Sun-Earth Connection Education Forum and High Energy Solar Spectroscopic Imager Mission are sponsoring the broadcast, which is being conducted by the Exploratorium, a science museum in San Francisco. Beginning at Land's End in the United Kingdom, correspondents for the Exploratorium will provide updates of the eclipse across Europe.

Japan's satellite switches target after fuel hitch

[TOKYO] Japan's Institute of Space and Astronautical Science announced last week that it has changed the target asteroid of MUSES-C — its third engineering test satellite — from Nereus to 1989ML, which is closer to Earth.

According to the institute, the four-year mission was re-routed after problems with the satellite's fuel capacity. The technical goals of MUSES-C, to be launched in July 2002, include the use of a solar-powered propulsion system and a high-speed re-entry through the Earth's atmosphere. It will also collect asteroid samples.

Deep Space 1 misses photos of asteroid

[WASHINGTON] The Deep Space 1 spacecraft, launched by the US space agency NASA, successfully flew close to the asteroid Braille last week, but failed to capture detailed pictures of the mile-wide object.

Early this week, engineers at NASA's Jet Propulsion Laboratory were still examining data from the fly-by, which was intended as a difficult test of the spacecraft's ability to plot its own course and plan photographic sequences of a rapidly moving target (see *Nature* **400**, 392; 1999). The autonomous navigation system apparently did not lock the cameras on to the asteroid during closest approach, perhaps because Braille was dimmer than expected.

Cancer supercomputer raises capacity 50 times

[WASHINGTON] The US National Cancer Institute has signed a \$6.5 million agreement to lease a Cray SV1 supercomputer, replacing a 1991 machine with a computer that has 48 times more capacity. The machine is being leased for three years from SGI of Mountain View, California (formerly Silicon Graphics).

It will be operated at the institute's Advanced Biomedical Computer Center in

Frederick, Maryland. Stan Burt, the centre's director, says that the Cray SV1 "represents a significant resource for the entire biological community".

The centre, which has an operating budget of \$3.9 million, is the only supercomputing operation in the United States dedicated solely to biological research.

Web puts jargon into the dictionary faster

[LONDON] The publisher of the *Oxford English Dictionary* (OED) is issuing a broad appeal to residents of English-speaking countries for documentary evidence of new words. In the past 18 months, the OED has appealed for evidence on words such as space junk, termiting, tipitiwitchet, asteroid field and alkalimetrically.

For the first time in its history, the dictionary is being completely revised. The project will cost £35 million (US\$55 million) and be completed by 2010. Next year the OED will go online, incorporating at least 1,000 new and revised entries every three months.

The OED regularly appeals for information on words or phrases, where documentary evidence is incomplete. Evidence can be submitted at: <http://www.oed.com>.

More haste, less science?

Sir — One widely recognized trend in scientific research is the development of the 'publish or perish' academic environment. A second trend is the increasing complexity and technical sophistication of science. Third, the amount of science that is being published is increasing rapidly, and scientists complain about the difficulty of keeping abreast of the literature. One consequence of these trends has not, to my knowledge, been adequately addressed: namely, how has the accuracy of the literature been affected by these pressures?

I sampled a leading international journal and quantified the number of original articles that contained at least one error. I read correspondence, errata and technical articles in one issue per month (usually the first week's issue), and recorded all published errors by authors or editors. I did not count disputes between authors and correspondents unless authors admitted that critics had identified a legitimate mistake.

I classified errors as: 'trivial' (typographical, grammatical or production errors of little or no impact on interpretation of the paper); 'errors of scholarship' (failures to acknowledge priority, mistaken claims of novelty, misuse of terminology, and improper or missing citations); and 'technical errors' (author- and journal-generated errors that could cause confusion, such as errors of fact or omission, statistical errors, chemicals misidentified, errors in mathematical equations or DNA sequences, figure legends transposed, use of improper methodology, data misinterpreted, and, in extreme cases, the retraction of parts or all of a paper).

I also attempted to distinguish whether the error originated with the authors or with the journal, although this was possible in only 81% of the papers. Finally, I counted the total number of articles and reports in each issue of the journal to estimate error

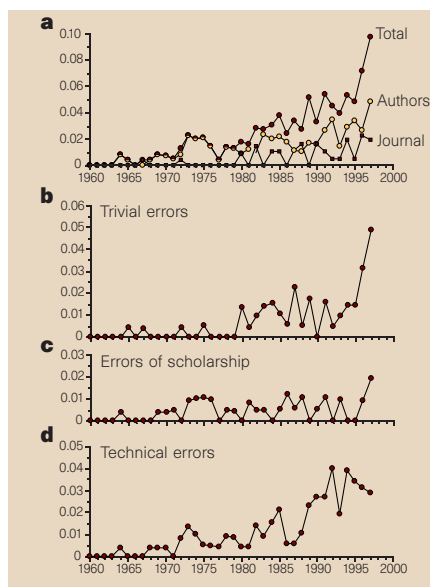


Figure 1 The yearly error rate in articles and reports in a major scientific journal from 1960 to 1997. Based on a sample of 12 issues per annum. a, Total rate and the rates judged to be due either to authors or to the journal. b–d, Errors broken down by type and severity.

rates (total number of papers containing errors per year/total number of papers that year). I sampled from 1960 (when professorial life was more leisurely, according to my old PhD supervisor) to 1997. Because I depended on authors, correspondents or editorial staff to identify errors in papers, my estimates of error rates are almost certainly conservative.

I found 183 papers with errors: 56 (31%) trivial, 35 (19%) errors of scholarship, and 92 (50%) technical errors. The error rate has been increasing since the early 1960s, when there were few reported errors (Fig. 1). By the 1970s, errors began to appear regularly, almost all of which were, as far as I could judge, the authors'

responsibility. But by the 1980s and 1990s, both authors and the journal have contributed to the error rate, although authors remain responsible for most.

Although a third of errors can be considered trivial, the number of such errors is increasing rapidly, suggesting that production standards are becoming more difficult to maintain. There is no strong temporal trend in errors of scholarship, at least since the 1970s, suggesting that authors have not yet lost track of the literature. By contrast, technical errors have been increasing.

What factors are likely to be contributing to this state of affairs? First, authors may be less careful during manuscript preparation owing to increasing career pressures to publish. Second, reviewers and editors may also be less thorough. Third, the increasing complexity of science may mean that more errors escape notice until papers are published. Fourth, journals may increasingly be contributing to errors in the literature. An alternative line of reasoning is to assume that the innate error rate has not changed but that we are more likely to criticize others' work than in the past, or that authors are generally more willing to accept criticism than in the past.

I believe that the observed increase is real, and that the academic environment is generating sources of error that are getting more severe. I doubt that this problem will be solved unless there are significant changes in the criteria by which many of us are evaluated by our universities. We might also ask ourselves about the consequences for scientific progress of the apparent increase in the publication error rate.

Bradford A. Hawkins

*Department of Ecology and Evolutionary Biology,
University of California,
Irvine, California 92697, USA*

Tale of an intrepid duo unearthed

Sir — In your 100 Years Ago column you reprinted a report on archaeological digs carried out in Egypt in the 1890s by a Miss Benson and a Miss Gourlay (*Nature* 398, 469; 1999). The Miss Benson referred to was Margaret ("Maggie") Benson, daughter of an Archbishop of Canterbury who had died a couple of years before the article's original appearance. One of her brothers, E. F. Benson (already a best-selling author with his novel *Dodo*, and later to write the

popular *Mapp and Lucia* novels), spent time each winter assisting at these excavations.

The ladies unearthed more than 200 statues, most of which went to Egyptian museums, apart from a very few items (undoubtedly those considered the least valuable and interesting), which the Egyptian government presented to their finders in recognition of their considerable contribution to the history of the XXVth and XXVIth dynasties.

It is galling that the *Nature* writer would never know that just a handful of what he so disparagingly dismissed as "no very startling discoveries" were to change hands

for huge sums of money in the twentieth century. In 1972, six of Miss Benson's statues were sold in London for a total of £114,000. And the head of Amun, which sold for 17,000 guineas in 1972, was the top lot at a Christie's auction in 1991, fetching more than £0.5 million.

The reporter's patronizing tone towards women has to be seen in the context of the day, but doesn't it bring home the attitudes with which intelligent and enterprising women had to contend a century ago?

Cynthia Reavell

*The Tilling Society,
5 Friars Bank, Pett Road, Guestling,
East Sussex TN35 4ET, UK*

How to restore public trust in science

The relationship between the scientific community and the general public has never been worse in living memory. The commercialization of research is largely responsible, but scientists can still act on the problem.

Benny Haerlin and Doug Parr

The scientific community has a credibility problem. In a recent survey, for example, 26% of European citizens named environmental organizations when asked whom they trusted most to tell the truth about genetically modified (GM) crops¹. Only 6% named universities, 4% national public authorities and 1% industry.

Such dwindling credibility of scientific institutions has been attributed to the disaster over bovine spongiform encephalopathy. But previous events had followed similar patterns: the debate about CFCs and the ozone hole in the 1980s, for example, the health and environmental assessment of toxic substances, global warming, and even smoking. The role of corporate scientists in these cases has not been admirable, and the attitude of industry and scientific institutions, in demanding conclusive proof to justify preventative action, has rightly not reassured the public about scientists' trustworthiness.

Such experiences led to the development of the 'precautionary principle' by legislators around the world to shift the burden of proof from protectors to perpetrators. It is unfortunate that many use 'sound science' as a battle cry against this principle, and erroneously dismiss environmental and ethical concerns as 'anti-scientific'.

Sound science is about the best possible way to answer a given question; to present with rigour the certainties and uncertainties of knowledge, and the assumptions underlying certain conclusions. But, crucially, it is not a method for deciding which questions should be posed, or for determining the acceptable risks and desirable benefits of technologies.

There are no clear answers to many of the 'big picture' scientific questions asked by the public, in many cases because we lack the knowledge, but in others because arbitration between different answers is beyond scientific competence. When asked about impacts on diversity and evolution, the shape of future agricultural landscapes, and the changing perception of food and health, honest scientists will frequently have to answer: 'We don't know', 'We cannot know' or 'These are our guesses'. Such honest answers could help a great deal to raise credibility and to return the responsibility for decision-making from corporate interests to its rightful place: public bodies.

Arguing that science is the sole arbiter of policy action undermines trust in the concept of scientific analysis. The main culprits in the devaluing of scientific authority are not nec-

Codes of conduct should be extended to make companies and scientists communicate information on environmental and health impacts of products

essarily scientists themselves but corporations and politicians, keen to rely on the illusory picture of authoritative scientific arbitrators. Scientists are no longer perceived exclusively as guardians of objective truth, but also as smart promoters of their own interests in a media-driven marketplace.

This changing role can be seen in molecular biology, where the line between fundamental science and applied technology appears particularly thin, and where corporate funding has become the driving force of research. Privatization of science has become official policy in all industrialized nations since the early 1980s. Not only has public spending on research in Europe, particularly in biology and agriculture, been dwindling over the past decade but, as Richard Strohmman bluntly observed, "academic biologists and corporate researchers have become indistinguishable, and special awards are given for collaborations between these two sectors for behaviour that used to be cited as a conflict of interest"².

Integration of academic research into the market, however innovative, demands a price on the role and credibility of scientists. Critics should first address governments that believe that the market should be their prime adviser on science policy. Not only is it unfair to blame scientists for losing their independence, but it stifles an urgently needed open and honest dialogue between scientists and the public.

One prerequisite for such a dialogue might be a reliable scheme of transparency on patents, financial interests and corporate affiliations. The first laudable steps being taken by journals and public authorities must lead to clear rules that can be checked by the public. Codes of conduct should be extended to compel companies and institutions to communicate information on environmental and health impacts of products, and to oblige individual scientists to communicate relevant findings. Such an analogy to the Hippocratic oath would strengthen the position and responsibility of scientists within corporate and institutional systems.

Beyond scientific independence and conflicts of interest, worries stem from an explosive increase in available data and specialization, and the likelihood is of increasing fragmentation of scientific knowledge and perception. Failures of science to predict negative outcomes seem to arise when a reductionist method encounters situations of high complexity. Interdisciplinary and holistic understanding about highly complex issues will not come from individual scientists, but will require entirely new and innovative approaches.

Who cares about this 'big picture', and who is responsible for the integration of knowledge? Corporations see little use in investing in integrated scientific concepts: their integrating forces are product development and marketing. Where are the scientific authorities and the editorials in journals challenging public and corporate research strategies and perspectives? Where are the scientific authorities to insist that the identification and patenting of individual genes of interest are not good enough to develop sound scientific understanding? Who dares to say that further specialization is a recipe for disasters?

Inevitably, the framing of these questions is affected by value judgements. The UK Royal Commission on Environmental Pollution advises that decisions on environmental issues "must be informed by an understanding of people's values", and that "The public should be involved in the formulation of strategies, rather than merely being consulted on drafted proposals."³

In the case of genetically engineered crops, surveys in Europe have consistently shown broad rejection of GM food. The governmental and corporate reaction has been to spend millions to 'educate' people about the perceived benefits of an 'inevitable' technology. The result has been greater understanding of the technology combined with firmer rejection of it. Instead of rethinking their research and development strategies and looking at the alternatives, most companies and governments still treat public acceptance as just an additional challenge to be overcome by asserting the safety of the technology. They are out of touch with the values of society, and that cannot be overcome by means of any scientific risk assessment. □

Benny Haerlin is at Greenpeace International, Chausseestrasse 131, D-10115 Berlin, Germany.

Doug Parr is at Greenpeace UK, Canonbury Villas, London N1 2PN, UK.

1. European Commission DG XII Eurobarometer 46.1, 77–78 (1997).

2. Strohmman, R. *Nature Biotechnol.* 15, 194–200 (1997).

3. Royal Commission on Environmental Pollution *Setting Environmental Standards* (HMSO, London, 1998).

Bollworms, genes and ecologists

M. J. Crawley

A big fuss is likely to be made over new ecological research results on a genetically modified crop. But what are the scientific questions about GM crops that ecologists are attempting to tackle?

In a brief communication on page 519 of this issue¹, Liu *et al.* describe research showing that if an insect pest is fed on a genetically modified (GM) crop, its developmental timing alters. The pest concerned is the pink bollworm (*Pectinophora gossypiella*); the crop is cotton, which has been genetically engineered to include genes encoding insect toxins from the bacterium *Bacillus thuringiensis* (*Bt*). Laboratory strains of larvae resistant to the toxin take longer to develop than non-resistant larvae raised on unmodified cotton.

The main implication of the study is that it calls into question the 'refuge strategy' of reducing or delaying the evolution of resistance to *Bt* toxin. The idea behind the strategy is that toxin-resistant pests will mate, and thus mix genes, with susceptible forms of the same species raised in adjacent 'refuges' of unmodified crops. The strategy might be expected to work because resistance is usually a genetically recessive trait. But if insects develop to sexual maturity at different speeds, interbreeding is less likely or impossible, and so may well speed the development of resistance and reduce the benefits of the GM crop.

Only a few weeks ago we had Losey and colleagues' controversial study² of *Bt* pollen and monarch butterfly caterpillars, which showed that exposure to a dusting of *Bt*-expressing maize pollen on their wild host plants could increase caterpillar death rate. This further finding¹ will add to the debate over GM crops. The immediate issues are whether the results of these small-scale studies should be published, and how much store should be set by them. The answers are yes they should be published, for they constitute carefully done science that gives pointers towards issues that require further scientific investigation. But no, in policy terms at least, little can be read into them: both are laboratory studies; they were carried out over comparatively short periods of time and under a specific set of conditions; and they do not address the issues over the entire life cycle of the insects. So larger, longer-term trials are necessary³.

The broader context to this debate has been discussed in a Briefing article in this journal⁴, and the reasons for the debate's especial intensity in Britain are outlined in Box 1. Several different issues — ethical, economic and health-related — have become

entangled and this greatly complicates any consideration of the ecological questions that stem from GM crops. The ecologist's remit is actually clear: to understand the distribution and abundance of organisms by studying the way that biological interactions (such as competition and predation) are affected by abiotic factors (such as weather or soil type). In the context of GM plants, this requires a detailed consideration of how the birth, death and dispersal rates of the organism could be affected by the possession of the transgene; and of how indirect effects — acting through the plant's nutrients, and its enemies and competitors, mutualists or decomposers (call them interacting organisms) — might affect its demography.

So the contribution of ecologists has to do with predicting the ecological behaviour of GM organisms once they are released into the environment. We need to know whether they will increase in abundance, and whether they might affect the distribution and abundance of non-target species. We also need to understand the indirect ecological effects that could arise as a consequence of adopting a GM-based system of production. For example, the husbandry of GM crops might mean alterations in the frequency and timing of applications of pesticides and fertilizers, or alterations to crop rotation and cultivation regimes.

Early field experiments on ecological risk assessment established that there was nothing

fundamentally different about GM crop plants. For instance, a comparison of the population dynamics of GM and conventional oilseed rape, maize, potato and sugar beet, carried out by myself and colleagues in a range of habitats, found that in no case were the GM plants more invasive or more persistent than their conventional counterparts⁵.

But what about outbreeding and 'gene escape'? Work on pollen dispersal and hybridization showed that there was the potential for transfer of herbicide resistance — another feature that can be 'engineered in' — to weedy or potentially weedy crop relatives⁶. However it is not gene movement as such that is the real issue. The important point is whether or not the product of cross-pollination poses a threat. If the plant that grows from the hybrid seed is not a problem, then the distance of pollen movement is irrelevant. But if the hybrid plant is a problem, say because it is a herbicide-resistant weed, then the genetic modification should not be introduced in the first place. Distance will not protect us; if cross-pollination can occur, it will. A bee that gets on a train could deliver its cargo of pollen to far-flung places.

At first sight, the results described by Liu *et al.*¹ and Losey *et al.*² seem straightforward. But interactions and nonlinearities are such an essential part of most ecological systems that they cannot be taken at face value. In the case of the monarch butterflies², increased mortality at one stage of the life cycle does not necessarily mean reduced population size in the long run, because there may be compensating density-dependence acting at some other stage of the life cycle. Likewise, an increase in the frequency of mating between bollworms with *Bt*-resistant genes may not lead to an overall increase in resistant-gene frequency if resistant bollworms suffer a higher death rate (for example higher over-winter mortality may result from their delay in maturation). It is essential, therefore, to embed findings such as these in models of population dynamics that integrate the entire life cycle.



Figure 1 Cotton pest — the pink bollworm, seen here in a damaged cotton boll.

HOLT STUDIOS/NIGEL CATTILIN

Box 1: GM crops and the British public

Why is opposition to GM crops so much more intense in Britain than, say, in the United States or Germany? It seems to me that there are three main reasons: the BSE crisis involving 'mad cow' disease; a growing disaffection with modern agriculture; and scepticism about 'big agribusiness'.

The BSE crisis raised various concerns (most notably fears over the large-scale transmission of neurodegenerative disease to humans through contaminated beef) that were powerful enough on their own but together made a deep impression on the national psyche. It was yet another food scare, following hard on the heels of *Salmonella* in eggs, and deepened the public's mistrust of scientists, politicians and government officials.

Beyond BSE, there is dissatisfaction with

modern farming methods and their impact on the countryside. Many people also regard the use of antibiotics and growth hormones in animal production as a threat to human health, and see factory farming and the cloning of farm animals as unethical.

Finally, elements of the public are suspicious of multinational agrochemical companies, seeing them as controlling all of the means of production. Talk of introducing designer genes to prevent farmers using home-grown seed to plant next year's crop only added to a sense of injustice.

In 1998, adding to this mix of fears and anger, came the widely reported import of products containing unlabelled GM soya from the United States. Attempts at media

management backfired, and pronouncements that GM foods would liberate the Third World from starvation were met with scorn – even the least sceptical newspaper reader realizes that GM crops are being developed for profit not philanthropy.

Another, very real economic concern, is driven by consumer choice. The organic farmer who cannot sell his crop because it has been dusted with wind-borne pollen from his neighbour's GM maize has a genuine problem.

Those involved in ecological risk assessment of GM organisms have to disentangle these concerns, and to show which are tractable scientifically. The task is made all the more difficult by being set within a potent mixture of opinion, both rational and irrational. **M. J. C.**

The kind of case-by-case experimental approach^{1,2} adopted in GM risk assessment research to date is vital, because it demonstrates the kind of things that can happen. But it is not the answer to our current dilemma: we need to predict the environmental consequences of a fundamental change in agricultural practice. It would be impossibly expensive to test every GM crop under every combination of environmental conditions. Likewise, no single experiment could be carried out at a sufficiently large spatial scale, or over a sufficiently long period of time, to convince the sceptics that all the issues had been addressed. Those who object to GM technology will always protest that scientists cannot guarantee that nothing could go wrong. This is true, of course — you cannot prove a negative, and absence of evidence is not evidence of absence.

Nonetheless, by concentrating on mechanisms (the direct and indirect effects on interacting organisms) and on how, exactly, GM individuals might differ from non-GM individuals in these respects, ecologists are in a position to carry out sound risk assessment procedures. The necessary theory is reasonably well developed³; we know what to measure, and how to measure it.

There are two interrelated parts to the process. Sharply focused experiments carried out under field conditions involve manipulation of the resources and interacting organisms of the GM organism and its unmodified counterpart⁵. The experimental results are then used to estimate the parameter values of a whole-life-cycle model. The model consists of a set of coupled equations which describe how change in the abundance of the GM organism affects and is affected by change in the interacting organisms. It is a simplified caricature of reality that aims to demonstrate the quantitative consequences of changes in assumptions about how environmental variability interacts with the species' nonlinear dynamics. As Einstein said: "A model should be as simple as possible. But no simpler". Analysis and numerical simulation with the model suggest how introduction of the GM organisms might influence whole-system dynamics, and generate new hypotheses for testing. These in turn lead to further field experiments and refinement of the model.

We know how to do the science. But it is by no means clear that, in Britain at least, the public mood is ready to be swayed by scientific results, however well replicated and

carefully controlled the experiments. Opposition to GM crops has become so militant (witness the recent destruction of transgenic poplar trees and GM maize crops) that the British government is considering changing the law in order to protect experimental sites by keeping their locations secret.

It is simply not clear at present whether the overall environmental impact of GM technology will be positive or negative. It could be beneficial if use of GM organisms leads to much lower use of pesticides, fertilizers or antibiotics. Alternatively, growing GM crops could exacerbate current trends of declining biodiversity, by encouraging farmers to keep farmland free of anything but crop plants for even more of the year. Field margins and roadsides that are currently such an important refuge for beneficial insects and other wildlife, may be seriously damaged. Only time will tell.

This leaves policy makers with a real dilemma. The long-term environmental impact of the new technology is uncertain, but they do not want to hamper the development and competitiveness of national agro-industry or biotechnology. On the other hand, there is real opposition to GM foods which is coupled to an increasing interest in organic farming. This is dismissed as 'all muck and mystery' by its opponents and, indeed, many aspects of organic farming are hopelessly unrealistic in a competitive global market. Policy decisions, therefore, will not depend only on ecological issues — crucially, they are likely to hinge on assessments of the risks and benefits of conventional, versus GM, versus organic farming.

The studies reported by Liu *et al.*¹ and Losey *et al.*² provide a glimpse of some of the pieces of the ecological jigsaw. But we shall not understand the big picture until the pieces are embedded in models of the entire life-cycles of the insect pest and its food plant, which also take account of the dynamics of the main interacting organisms. Until that is done we cannot gauge the long-term effects of genetic engineering on gene frequencies or population size in the field. In the meantime, ecologists must do the best they can in providing an assessment of risks and benefits. Such work may not change public perceptions, because these are bound up with so many other emotive issues. But even if science cannot offer certainty it can strive to offer objectivity. □

M. J. Crawley is in the Department of Biology, Imperial College, Silwood Park, Ascot SL5 7PY, UK. e-mail: m.crawley@ic.ac.uk

1. Liu, Y.-B., Tabashnik, B. E., Dennehy, T. J., Patin, A. L. & Bartlett, A. C. *Nature* **400**, 519 (1999).
2. Losey, J. E., Rayor, L. S. & Carter, M. E. *Nature* **399**, 214 (1999).
3. Firbank, L. G. *et al.* *Nature* **399**, 727–728 (1999).
4. Butler, D. *et al.* *Nature* **398**, 651–656 (1999).
5. Crawley, M. J. *et al.* *Nature* **363**, 620–623 (1993).
6. Brown, J. & Brown, A. P. *Ann. Appl. Biol.* **129**, 513–522 (1996).
7. Gurney, W. S. C. & Nisbet, R. M. *Ecological Dynamics* (Oxford Univ. Press, 1998).

Palaeoecology

A breath of fresh air

D. E. Canfield

The Earth has not always been bathed in an oxygen-rich atmosphere. Models of the early atmosphere suggest that it was oxygen-free, and contained chemically reduced compounds such as carbon monoxide, hydrogen gas and possibly methane¹. Oxygen could not accumulate until microorganisms that produced oxygen by photosynthesis, known as cyanobacteria, evolved (Fig. 1). This was a defining moment in Earth history. The availability of oxygen allowed the evolution of aerobic organisms, and forever changed the chemistry of the planet's surface environment. When, then, did cyanobacteria evolve? On page 554 of this issue, Summons *et al.*² describe molecular fossils derived from the cell walls of cyanobacteria found in 2.5-billion-year (Gyr)-old sedimentary rocks from the Mount McRae shale in Western Australia. This is the earliest direct evidence for cyanobacterial activity.

Molecular fossils (biomarkers) are organic compounds that were derived from living organisms, and are found in sedimentary rocks rich in organic compounds. To be useful, biomarkers must retain some resemblance to the original biological molecules, even after the long history of decomposition and alteration that accompanies their burial and sedimentary rock formation. A biomarker must also be diagnostic — that is, it must be indicative of a certain group of organisms. In cyanobacteria, a class of bacteriohopanepolyols (BHP) serve to make the cell-wall rigid and are frequently methylated at the C-2 position (2-methyl-BHP). These 2-methyl-BHP compounds have not yet been found in quantitatively significant amounts in any other group of organisms, and, as far as we know, are uniquely diagnostic of cyanobacteria. Luckily, through burial and thermal maturation, 2-methyl-BHP is converted to easily recognizable, and still diagnostic, 2-methylhopanes. A constant concern with old sedimentary rocks of Precambrian age is that biomarker signatures have arisen from a contaminating source, such as remobilized organic compounds from nearby, younger sedimentary strata. Summons and colleagues² have ruled out this possibility, leaving us with the fingerprint of cyanobacterial activity.

Other indirect evidence suggests that cyanobacteria may be much older than this. The isotopic composition of the organic carbon preserved in sedimentary rocks shows a decided shift to values that are very ¹³C-depleted around 2.7–2.8 Gyr ago³. This isotope trend is believed to show the incorporation of highly ¹³C-depleted biomass from

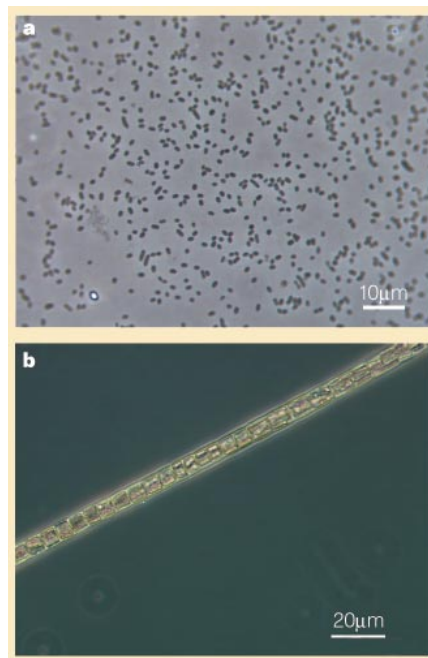


Figure 1 Two of the most abundant modern open-ocean cyanobacteria. a, *Synechococcus*, species of which are unicellular, and are important primary producers in the modern ocean. b, A filament of *Trichodesmium*. These species also contribute to primary production and are important fixers of atmospheric nitrogen. Organisms such as these may resemble early cyanobacteria, whose 2.5 billion-year-old biomarker record is reported by Summons and colleagues².

methanotropic bacteria into the sedimentary carbon record. Methanotropic bacteria oxidize methane, and the shift to highly ¹³C-depleted values could, therefore, indicate the emergence of oxygen-producing cyanobacteria. Furthermore, geological and geochemical considerations point to the involvement of cyanobacteria in the formation of stromatolites (believed to be fossilized, layered microbial communities) in lakes 2.7 Gyr ago⁴. Older still, bacterial microfossils miraculously preserved in 3.47-Gyr-old sedimentary rocks bear a striking resemblance to modern cyanobacteria, although morphology alone is not a definitive indicator of bacterial affiliation⁵.

The biomarker studies of Summons and colleagues² make it nearly certain that cyanobacteria evolved well before 2.5 Gyr ago, although the precise timing is still elusive. Nevertheless, cyanobacteria evolved on an anoxic Earth that surprisingly remained that way until around 2.2 Gyr ago. It is not until then that the geological record provides compelling evidence for widespread oxidation weathering on land — a sure sign of oxy-

gen accumulation in the atmosphere⁶. Why should the evolution of cyanobacteria and the large-scale expression of its metabolism — that is, the oxidation of the Earth's surface environment — be so far separated in time? There are several possible explanations. The accumulation of oxygen at the Earth's surface requires the burial and preservation into sediments of organic carbon ultimately produced by oxygenic photosynthesis. If the continental margins, where most organic carbon burial occurs, were much smaller on the early Earth, then less organic carbon might have been buried, liberating less oxygen to the surface⁷. Low concentrations of atmospheric oxygen might also have been maintained by a high flux of reduced chemical species from the Earth's mantle, creating a large demand for oxygen⁶.

There may be yet another possibility. Both biochemical considerations⁸ and ribosomal RNA evolutionary histories⁹ show that cyanobacteria were not the first photosynthetic organisms. Rather, anoxygenic photosynthetic organisms evolved first, and these use reduced chemical species such as hydrogen sulphide and ferrous iron as electron donors¹⁰, producing oxidized sulphur species and ferric iron oxides. At the time of cyanobacterial evolution, deep-ocean waters contained appreciable concentrations of ferrous iron, as evidenced by the widespread occurrence of banded-iron formations in Archaean (older than 2.5 Gyr) sedimentary rocks. Therefore cyanobacteria may well have evolved in an ecosystem with a well-established anoxygenic photosynthetic population which used ferrous iron and nutrients from the deep ocean, produced iron oxides and perhaps contributed to the genesis of banded-iron formations¹⁰.

Oxygenic photosynthesis in the open ocean would have been limited and controlled by the supply of nutrients escaping from the anoxygenic photosynthetic population. A better refuge for early cyanobacteria may have been shallow coastal areas free of ferrous iron from the deep ocean. In this scheme of events, photosynthetic oxygen production (both by cyanobacteria and later eukaryotic organisms) could not become established until the supply of reduced species in the deep ocean became limiting. This may have occurred in stages, with the final oxygenation of the deep ocean occurring late in the Precambrian, perhaps only 700–800 million years ago¹¹.

Many key events in bacterial evolution and the oxidation of the Earth's surface are still unknown. Biomarker studies such as that of Summons and colleagues² hold great promise in unravelling the timing of evolutionary milestones, and in deducing the structure of ancient ecosystems. □

D. E. Canfield is at the Institute of Biology, Odense University, Campusvej 55, 5230 Odense M, Denmark. e-mail: dec@biology.ou.dk

1. Kasting, J. R. *Science* **259**, 920–926 (1993).
2. Summons, J., Jahnke, L. L., Hope, J. M. & Logan, G. A. *Nature* **400**, 554–557 (1999).
3. Hayes, J. M. in *Early Life on Earth* (ed. Bengtson, S.) 220–236 (Columbia Univ. Press, New York, 1994).
4. Buick, R. *Science* **255**, 74–77 (1992).
5. Schopf, J. W. *Science* **260**, 640–646 (1993).
6. Holland, H. D. in *Early Life on Earth* (ed. Bengtson, S.) 237–244 (Columbia Univ. Press, New York, 1994).
7. Des Marais, D. J. *Org. Geochem.* **27**, 185–193 (1997).
8. Blankenship, R. E. *et al. Trends Biochem. Sci.* **23**, 94–97 (1998).
9. Woese, C. R. *Microbiol. Rev.* **51**, 221–271 (1987).
10. Widdel, F. *et al. Nature* **362**, 834–835 (1993).
11. Canfield, D. E. *Nature* **396**, 450–453 (1998).

Quantum optics

Energy transfer under control

William L. Barnes and Piers Andrew

Life on Earth is ultimately sustained by photosynthesis. This phenomenon makes use of perhaps the most fundamental example of a process known as excitation-energy transfer¹. Solar energy is harvested by 'antenna' molecules that readily absorb photons of light, and is then transferred with great efficiency to a distant reaction site for use in metabolic processes. This transfer mechanism has been widely studied; it has even proved possible to mimic this aspect of photosynthesis in artificial light-harvesting structures based on molecular films². Adopting this energy-transfer process in optoelectronics is an attractive idea that allows us to tailor the characteristics of a new generation of devices, such as polymer light-emitting diodes³ and lasers⁴.

Developments in quantum optics have shown that the interaction of light and matter can be controlled by manipulating the local optical environment⁵. Microcavities and photonic bandgap materials have been used to control spontaneous emission, in which an excited molecule gives up its energy as light. Can the same control be exerted over the energy-transfer process? A report in *Physical Review Letters*⁶ by Hopmeier *et al.* indicates that such control may indeed be possible, and could have far-reaching implications.

Excitation-energy transfer takes place between a donor molecule, initially in an excited state, and an acceptor molecule, usually in its ground state (Fig. 1). Provided there is sufficient overlap between the emission spectrum of the donor and the absorption spectrum of the acceptor, energy may be transferred. If the donor and the acceptor are far apart then the transfer is radiative — the donor emits a photon that is subsequently absorbed by the acceptor. For smaller separations, less than the wavelength of the light involved, the dipole moment associated with the excited donor may induce a dipole moment in the acceptor. The resulting dipole–dipole interaction leads to a non-radiative transfer of energy known as Oppenheimer, or Förster^{7,8}, transfer.

By placing an excited molecule between two closely spaced mirrors to form a microcavity, the process of spontaneous emission may be controlled. The boundary conditions imposed by a microcavity lead to resonant

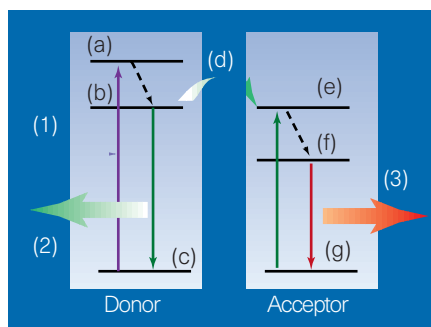


Figure 1 Energy-transfer scheme. Light (1) is absorbed by the donor, promoting it to an excited state (a) from which it relaxes non-radiatively to a metastable level (b). From here it may either return to the ground state (c) to produce a photon (2), or transfer its energy (d) and thus promote the acceptor to an excited state (e). The acceptor then relaxes (f) and may subsequently decay to the ground state (g) to produce a photon (3). The energy-transfer step (d) is the one that Hopmeier *et al.*⁶ are now able to control through the optical environment in which they place the donor and acceptor.

modes when the wavelength fits neatly within the cavity. If emission takes place at this resonant wavelength it will be enhanced. Several recent theoretical studies indicate that the energy-transfer process can be controlled in the same way as spontaneous emission by confining the molecules within, for example, a microcavity^{9,10} or a photonic bandgap material^{11,12}.

Hopmeier *et al.*⁶ have now studied the energy-transfer process in a well-defined microcavity consisting of a dielectric stack mirror on one side and a metallic mirror on the other. They studied energy transfer from a light-emitting polymer to an organic dye confined within this cavity. The polymer (donor) was excited with an ultraviolet laser, and emission from the dye (acceptor) was collected in the red part of the spectrum. By varying the thickness of the mixed polymer/dye layer separating the microcavity mirrors, they were able to sweep the resonance wavelength of the cavity all the way across the spectral region associated with energy transfer from polymer to dye. Using both intensity- and time-resolved measurements, they showed that a significant enhancement in emission from the dye



100 YEARS AGO

"To those who study the progress of exact science, the common spinning-top is a symbol of the labours and the perplexities of men who had successfully threaded the mazes of the planetary motions. The mathematicians of the last century, searching through nature for problems worthy of their analysis, found in the toy of their youth ample occupation for their highest mathematical powers. No illustration of astronomical precession can be devised more perfect than that presented by a properly balanced top, but yet the motion of rotation has intricacies far exceeding those of the theory of precession. Accordingly we find Euler and D'Alembert devoting their talent and their patience to the establishment of the laws of the rotation of solid bodies. Lagrange has incorporated his own analysis of the problem with his general treatment of mechanics; and since his time Poincaré has brought the subject under the power of a more searching analysis than that of the calculus, in which ideas take the place of symbols, and intelligible propositions supersede equations." (Maxwell – "Collected Works", I. p. 248). From *Nature* 3 August 1899.

50 YEARS AGO

Although much had been written and much had been done, nevertheless neither the Medical Research Council in Great Britain nor the National Research Council in the United States of America was in a position to give any definite answer when, at the beginning of the War, their respective Governments asked for information about the amounts of certain nutrients required by the human body. ... The intention was to give a group of volunteers "a diet virtually devoid of vitamin A and carotene until unmistakable signs of deficiency appeared, and then to determine what dose of vitamin A or carotene was needed to ensure recovery to normal". ... Perhaps the most unexpected findings in the present instance were the "prolonged delay before the onset of nutritional changes" (eight months of deprivation produced no discernible indication of deficiency). ... Defective night-vision, with a raised rod-threshold and prolonged cone-rod transition-time, and a lowering of the plasma level of vitamin A proved to be the first definite signs of deficiency. From *Nature* 6 August 1949.

occurred when the wavelength of the cavity mode coincided with that of the transfer process.

The authors' samples contained a random distribution of donors and acceptors within the cavity region. As a consequence, both radiative transfer between widely separated donor-acceptor pairs and non-radiative transfer between neighbouring pairs is expected. They argue that the increased acceptor emission they saw resulted from the microcavity enhancement of the radiative energy-transfer process. So, their results strengthen previous findings involving the optical modes of liquid microdroplets¹³.

The potential applications of this work are many and varied — for example in the design of organic light-emitting devices — but our knowledge of the underlying physics remains incomplete. The question remains as to whether the non-radiative transfer process is affected in a similar fashion. Further work with well-defined donor-acceptor separations is needed to give a clearer picture of the different transfer mechanisms. Nonetheless, it looks increasingly likely that energy transfer modified in this way will extend our ability to control the interaction between light and matter. One of the most recent uses to which energy transfer has been

put is to control the wavelength and efficiency of solid-state semiconducting organic lasers⁴. In the future, modifying the transfer process in the manner suggested by Hopmeier *et al.*⁶ should improve the efficiency of these lasers. This approach might also be used to control energy transfer between quantum dots¹⁴, and could perhaps be used to develop efficient artificial photosynthetic devices. □

William L. Barnes and Piers Andrew are in the School of Physics, University of Exeter, Exeter EX4 4QL, UK.

e-mails: w.l.barnes@ex.ac.uk

p.andrew@ex.ac.uk

1. Hu, X. C. & Schulten, K. *Phys. Today* **50**, 28–34 (1997).
2. Yamazaki, I., Tamai, N. & Yamazaki, T. *J. Phys. Chem.* **94**, 516–525 (1990).
3. Baldo, M. A. *et al.* *Nature* **395**, 151–154 (1998).
4. Berggren, M. *et al.* *Nature* **389**, 466–469 (1997).
5. Milonni, P. *The Quantum Vacuum* (Academic, London, 1994).
6. Hopmeier, M. *et al.* *Phys. Rev. Lett.* **82**, 4118–4121 (1999).
7. Oppenheimer, J. R. *Phys. Rev.* **60**, 158 (1941).
8. Förster, Th. *Ann. Phys. (Leipzig)* **2**, 55–75 (1948).
9. Kobayashi, T. *et al.* *Phys. Rev. A* **52**, 2835–2846 (1995).
10. Agarwal, G. S. & Dutta Gupta, S. *Phys. Rev. A* **57**, 667–670 (1998).
11. John, S. & Quang, T. *Phys. Rev. A* **52**, 4083–4088 (1995).
12. Kurizki, G. & Genack, A. Z. *Phys. Rev. Lett.* **61**, 2269–2271 (1988).
13. Folan, L. M. *et al.* *Chem. Phys. Lett.* **118**, 322–327 (1985).
14. Kagan, C. R. *et al.* *Phys. Rev. B* **54**, 8633–8643 (1996).

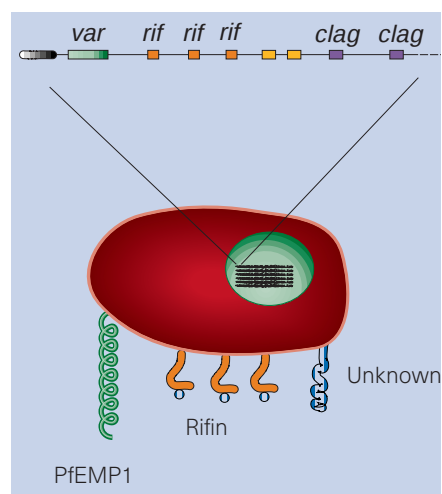


Figure 1 Red blood cell infected with *Plasmodium falciparum*. Antigens expressed at the surface of the blood cell are shown (bottom), as well as the end of chromosome 3 (top), where Bowman *et al.*¹ have identified the genes that encode these proteins. The antigens are highly variable, particularly those encoded by members of the *var* and *rif* gene families (PfEMP1 and rifin, respectively). There are roughly 50 *var* genes, although only one is expressed at a time^{11,12}. By contrast, several of the 200 or so *rif* gene products may be present on the surface of the same red blood cell. The *clag* gene family is smaller, and it is not yet clear whether the proteins encoded by these genes are expressed at the surface of the red blood cell⁷. However, parasites that lack the *clag* gene on chromosome 9 adhere poorly to host cells¹⁰. (Yellow boxes depict genes encoding proteins of unknown function.)

blood cells, where it is protected from most mechanisms of the immune system. But the parasite exports some of its own polypeptides to the surface of the red blood cell (Fig. 1), and here it is open to immune attack. To fool the immune system, it regularly exchanges these polypeptides ('antigenic variation')⁴. The polypeptides are encoded by members of the *var* (variable) and *rif* (repeated-interspersed family) gene families, which are found close to the ends (telomeres) of both chromosomes 2 and 3 — each telomere is flanked by one *var* gene and several members of the *rif* family. So far, nine *rif* genes (including two members of the *stevor* sub-family) have been identified in the subtelomeric regions of chromosome 3, and 17 members (including four *stevors*) have been found in the same regions of chromosome 2. This gives an estimated 200 *rif* genes per genome, making it the largest gene family in *P. falciparum*. Bowman and colleagues' results¹ make the picture even more complicated, as they have discovered more gene families — the *clag* (cytoadherence-linked asexual gene) and CTP (conserved telomeric protein) families, for example — in the subtelomeric region of chromosome 3.

Malaria

A blueprint of 'bad air'

Mats Wahlgren and Maria Teresa Bejarano

The 'bad air' agent — *malaria* in Italian — was fancied to be the cause of malaria far into the nineteenth century. At that time, Ross and Grassi discovered that mosquitoes transmit these microscopic protozoa. Just as this discovery was a big leap then, the next step today is to discern the parasite's blueprint; the genomic sequence of *Plasmodium falciparum*. Four hundred million people still suffer from malaria worldwide, and a million children die each year in Africa alone. There is no vaccine, anti-malarial drugs are failing, and many sufferers also develop tumours (such as Burkitt's lymphoma).

On page 532 of this issue, Bowman *et al.*¹ unveil the entire sequence of the *P. falciparum* chromosome 3. These authors have sequenced and analysed a million bases, which, along with the previously sequenced² chromosome 2, now gives us 7% of the total *P. falciparum* genome (roughly 30 megabase-pairs, Mb). This complex genome comprises 14 chromosomes, each equivalent in size to a complete bacterial genome (0.65 to 3.4 Mb), as well as mitochondrial- and plastid-like sequences. Some 215 protein-coding genes are found on chromosome 3, compared with 209 on chromosome 2, giving an estimated 6,500 genes in the whole genome.

Perplexing though the *P. falciparum* genome may seem, a certain organization has been found. Clusters of genes encoding proteins produced at similar stages in the parasite's life cycle, or proteins with similar functions, seem to be grouped in particular chromosomes or chromosomal locations. For example, in chromosome 2 there are many genes encoding proteins with high levels of the amino acid serine, and also genes encoding structures found on the surface of the parasite (merozoites)². Another example of gene clustering is the sexual-stage-specific genes, which are localized to chromosome 5 in a related malaria parasite. And a sequence that may be involved in separating a chromosome from its sister copy during cell division has been identified at roughly the centre of chromosomes 2 and 3. Known as the centromere, it consists of 2 kb of repeated adenine-thymine-rich sequences. However, the core of the *P. falciparum* centromere is shorter than those in other eukaryotes, and it remains to be shown whether it functions as a centromere. Such comparative analyses of chromosomes, particularly between different malarial parasites (all of which have 14 chromosomes)³, should help in understanding genome structure and evolution.

The malaria parasite develops inside red

Because several gene families are located close to each other, it is tempting to speculate that they are also co-expressed and co-regulated. The presence, in human sera, of antibodies that recognize the surface of an infected red blood cell has been shown^{5,6} to correlate with protection from malaria. Indeed, one polypeptide, PfEMP1 (*P. falciparum* erythrocyte membrane protein-1), is encoded by the *var* genes and is thought to be involved in antigenic variation. But the discovery of *rif*-gene products, rifins^{7,8} and stevors⁹, and the reactivity of rifins with human sera⁷, suggests that other molecules may also be involved. The rifins may, in fact, be a major source of antigenic variation at the surface of the red blood cell. Several rifins are sometimes expressed in the same red blood cell, and they also carry modifications that generate tiny variations between them⁷.

The PfEMP-1 protein is an adhesive molecule that mediates binding of the parasite-infected red blood cell to endothelial cells and to other red blood cells. The polypeptides encoded by the newly discovered *clag* gene family seem to have an accessory function in binding¹⁰. The *clag* genes have an interesting structure, with many exons ('coding' fragments, present in mature RNA) and introns (removed fragments). This structure is similar to that of genes in higher eukaryotes, and, although it is believed to be uncommon in *P. falciparum*, the structure has been found in 2.3% of the genes identified so far¹. This could even be an underestimate, because the structure is difficult to predict. Bowman *et al.* do not know whether the *clag* gene products are expressed on the surface of red blood cells, but other data¹⁰ indicate that they are biologically important.

Now that we have part of the genome sequence to hand, it is imperative that we use it to identify new drug targets, to understand the pathogenesis of malaria, and to help in constructing vaccines. However, the *P. falciparum* genome is more complicated than 14 chromosomes — it is plastic, recombinations occur, and it is modified at fertilization in the mosquito's gut. The genome may thereby be expanded, depending on whether the parasite is endemic in the locations studied. For example, if the transmission rate is high, the *var* and *rif* gene repertoires are thought to be restructured by frequent fusions of different haploid genomes. The opposite might be true in pockets of more restricted transmission. Self-fertilization has been found to produce genetically distinct parasites as well. So although the sequences of the chromosomes are a great step forward in understanding the parasite and the disease, there is still much to be done. 'Bad air' is still bad news. □

Mats Wahlgren and Maria Teresa Bejarano are at the Microbiology and Tumour Biology Centre,

Karolinska Institutet, and the Swedish Institute for Infectious Disease Control, Box 280, S-171 77 Stockholm, Sweden.
e-mail: mats.wahlgren@smi.ki.se

1. Bowman, S. *et al.* *Nature* **400**, 532–538 (1999).
2. Gardner, M. J. *et al.* *Science* **282**, 1126–1132 (1998).
3. Carlton, J. M.-R., Galinski, M. R., Barnwell, J. W. & Dame, J. B. *Mol. Biochem. Parasitol.* **101**, 22–32 (1999).
4. Wahlgren, M., Fernandez, V., Chen, Q., Svård, S. & Hagblom, P. *Cell* **96**, 603–606 (1999).

5. Carlson, J. *et al.* *Lancet* **336**, 1457–1460 (1990).
6. Bull, P. C. *et al.* *Nature Med.* **4**, 358–360 (1998).
7. Fernandez, V., Hommel, M., Chen, Q., Hagblom, P. & Wahlgren, M. *J. Exp. Med.* (in the press).
8. Keyes, S., Rowe, A., Kriek, N. & Newbold, C. A. *Proc. Natl Acad. Sci. USA* (submitted).
9. Chen, Q. *et al.* *Mol. Biochem. Parasitol.* **97**, 161–176 (1999).
10. Holt, D. C. *et al.* *Int. J. Parasitol.* (in the press).
11. Chen, Q. *et al.* *Nature* **394**, 392–395 (1998).
12. Scherf, A. *et al.* *EMBO J.* **17**, 5418–5426 (1998).

Engineering

Shark skin and other solutions

Philip Ball

What does nature have to offer the engineer? This was the gloriously rich question tackled at a meeting* on biomimetics and bionics held in June.

There are plenty of examples from engineering design that have, in retrospect, analogues in living systems: ball-and-socket joints, gear mechanisms, aerodynamic forms, laminated composite materials. But how much can nature be a guide in showing us solutions that would not have emerged anyway from sound engineering?

The classic example of inspiration from nature is Velcro, stimulated by the adherence of burred seed shells onto dog hair. Equally elegant is the way that studies of shark skin have led to drag-reducing coatings (D. Bechert, Technische Universität (TU) Berlin). The ribbed texture of the scales of a shark (Fig. 1), just visible under a magnifying glass, seems to provide aerodynamic efficiency relative to a smooth surface because of the way that the corrugations affect the viscous boundary layer of the water. A transparent plastic film with this same microscopic texture, ribs parallel to the direction of flow, helps to reduce aircraft drag by up to 8% — representing a fuel saving of about 1.5%.

Moreover, this textured film provides a self-cleaning function — the rough surface confers low wettability. Indeed, very highly textured surfaces can increase the contact angle of water to almost 180° even on otherwise hydrophilic materials (T. Onda *et al.* *Langmuir* **12**, 2125; 1996). An analogy in nature, although not the model in this case, is the lotus leaf, which has a hydrophobic, non-wettable surface produced by many micrometre-scale protrusions.

But 'learning from nature' is not necessarily about mimicry at a mechanistic level; perhaps a good algorithm is all the engineer needs. So, for example, genetic algorithms for minimization problems draw some inspiration from random mutation and natural selection, but no one would suggest that they are modelling evolution. Likewise, questions of stress minimization in structures can be addressed by an iterated finite-

element calculation that incorporates the basic principles of bone growth — adaptive growth, involving addition and removal in response to the stress field — without any attempt to capture the intricacies of bone-cell function (D. Reuschel, Forschungszen-

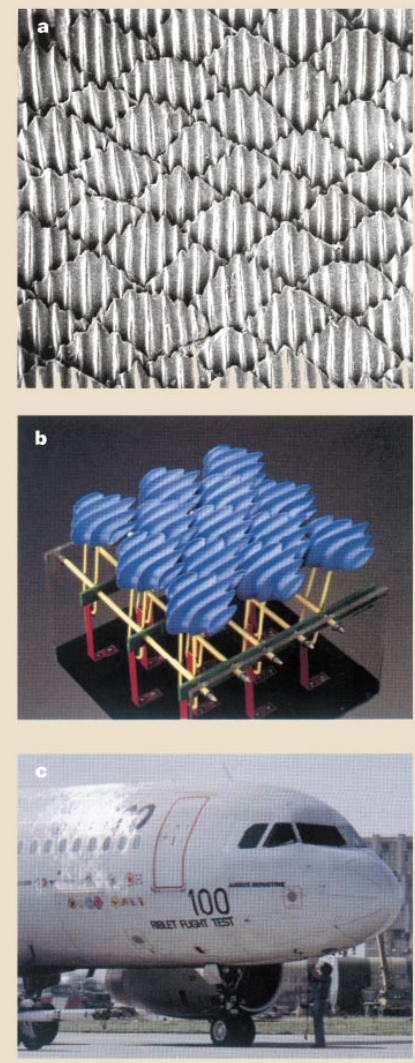


Figure 1 Corrugated coating for drag reduction. The riblets on shark skin (a) provided the inspiration for modelling studies of the drag reduction they confer (b), and eventually led to trials on an aircraft coated with a plastic film with this same microscopic texture (c).

*Inspiration from Nature, Institute for Advanced Study, Berlin, 9–11 June 1999.

trum Karlsruhe). This enables one to 'grow' a good engineering design out of a monolithic component, in which material has been removed where it serves no load-bearing function. The question is then whether the solution satisfies other engineering criteria, not least manufacturability.

Holes are a particularly irksome necessity for the engineer: rivets have to go somewhere. The problem is stress concentration — the same as enables a crack tip to propagate through a material at stresses far below the material's bulk fracture strength. Trees, however, have few problems with holes — or, equivalently, with inclusions such as knots — because these perturbations do not cleave the fibrous structure of the wood; rather, the fibres deform to take paths around the opening. It is a trick that can be usefully employed in fibrous composites (G. Jeronimidis, Univ. Reading) — if the holes can be incorporated during fabrication, rather than being drilled afterwards, stress concentrations can be partly compensated by the increased density of fibres. It is, however, more difficult to fix the delamination that tends to occur in layers parallel to the plane of the hole.

In this respect, perhaps more attention needs to be given to processing. After all, much that is attractive in natural materials synthesis — an aqueous environment, minimal waste and low energetic cost — is sacrificed if a natural microstructure is mimicked by high-temperature, high-vacuum processing (J. Vincent, Univ. Reading). (The flipside is slow growth speed.) In the much-studied case of spider silk, it seems that the remarkable strength derives not so much from the composition — the protein sequence — as from the spinning process, which seems to give an extraordinary woven texture to the fibrous strands (F. Vollrath, Univ. Aarhus). Synthesis of a genuine artificial silk might therefore demand much more than application of existing fibre-drawing methods to a synthetic peptide. Maybe only a micro-mechanical replication of the spinning gland will suffice.

Many natural materials also have the advantage of being active structures, not passive load-bearers, which is why muscle can exhibit infinite stiffness (no deformation for an increase in load). Muscle also begs the question of how far artificial actuators need to mimic the micro- and nanoscale structure to achieve comparable performance. The limited power density and/or small deformations available from current actuators — piezoelectric, magnetostrictive or polymer-gel-based — seem to be intrinsic to the actuating mechanisms, and perhaps only by building in some element of the interdigitated sarcomere structure of muscle can we hope to do better. A silicon-based micro-mechanical analogue is being developed (C. Schilling, TU Ilmenau), but there is no clear indication that silicon is the best material or

that the design remains practical when scaled up in this way.

Improved actuators are one of the primary requirements for robotics. In fact, Rodney Brooks at MIT has made a good case that robotics needs to put less emphasis on algorithms and more on their embodiment (R. Pfeifer, Univ. Zürich). That is, it is easy to over-design a robotic component (and thus to overload the processing requirements) when much of the control could instead be embodied in the properties of the materials used. For instance, our own limbs have far fewer degrees of freedom than could be incorporated in principle into those of a robot, and this simplifies the task of actually using them. Concentrating on embodiment can also remove preconceptions about morphological design: rather than trying to make an anthropomorphic robotic device indefinitely adaptable, it can be better to search for the ideal morphology (disposition of sensor and actuator units, say) for a given task environment. After all, natural systems are unlikely to suffer from over-design.

This touches on a recurring issue: is it valid to assume that natural forms are in some sense optimized? Natural selection has, of course, no short-term imperative towards

optimal design; all that matters are slight competitive advantages, achieved at minimal cost to the organism. But does this drive all organisms inevitably towards peaks in the fitness landscape? And does that necessarily make the natural design appropriate in engineering terms, in which 'cost' and 'competition' have very different meanings?

Certainly, it is striking that aerodynamic optimization of a torpedo-shaped body produces, at 'biologically relevant' Reynolds numbers, a solid body of rotation that has more in common with the shape of a swimming penguin than that of a submarine (R. Bannasch, TU Berlin). On the other hand, the use by some spiders of the GABA neurotransmitter as a hygroscopic coating for silk (the resulting droplets confer good capture properties by a kind of windlass effect) surely reflects less on any superior properties of GABA in this context and more on its simply being available (Vollrath).

But maybe all that really matters is good engineering judgement about whether a biological solution is a practical one. As Werner Nachtigall (Univ. des Saarlandes) put it, it is not necessary to use biology for inspiration, but may simply be wise to do so. □

Philip Ball is a consultant editor for Nature.
e-mail: p.ball@nature.com

Earth's productivity

Ghosts of biospheres past

Joe Berry

A variety of clever approaches has been developed to obtain proxy records of what the Earth was like in previous eras. Ocean cores, lake sediments, pollen, corals, palaeosols and the annual rings of trees have all been used to infer changes in climate and productivity¹. But these records are essentially local in coverage, and they need to be extended to regional or global scales. Such a planetary-scale analogue to the annual rings of trees would be a boon for Earth-system research, and the study by Luz *et al.*² on page 547 of this issue may provide just such information. Their work indicates that, by analysing the abundance of the three stable isotopes of oxygen — ¹⁶O, ¹⁷O and ¹⁸O — from air trapped in ice cores, we can obtain estimates of gross photosynthetic production of oxygen, integrated to a global scale and averaged over millennia.

Atmospheric oxygen is continuously recycled by biospheric processes — it is formed from water by photosynthetic organisms, and most of it is cycled back to water by aerobic respiration². It is estimated that the atmosphere now turns over in this way every 1,200 years³. Because the amount of O₂ in the atmosphere is roughly constant, the rate of turnover is inversely proportional

to the rate of total biological productivity summed for the whole planet. Luz *et al.*² use this variation in the turnover time of atmospheric O₂ to obtain proxy information on past global productivity from analyses of air extracted from glacial ice.

How did Luz and colleagues determine the turnover time of the atmosphere in the past? This is the ingenious part. Their method takes advantage of a little-known pathway of O₂ exchange that occurs in the stratosphere⁴ — the upper portion of the atmosphere, containing the ozone layer. The pathway involves photochemical reactions of ozone (O₃), and the isotope effect associated with this pathway is very different from that associated with biological cycling of O₂. Whereas the photochemical pathway has little or no preference for ¹⁷O over ¹⁸O, biological cycling shows a differential fractionation that depends on the mass of the isotopes⁵. Although there is no net loss of O₂ by this process, exchange reactions result in a net transfer of ¹⁷O from O₂ to CO₂, and finally to water (Fig. 1, overleaf). The authors propose that the stratospheric cycle would cause an isotopic anomaly in atmospheric O₂ (it should contain less ¹⁷O than expected), and that the size of this anomaly should depend on the relative rates of biological and

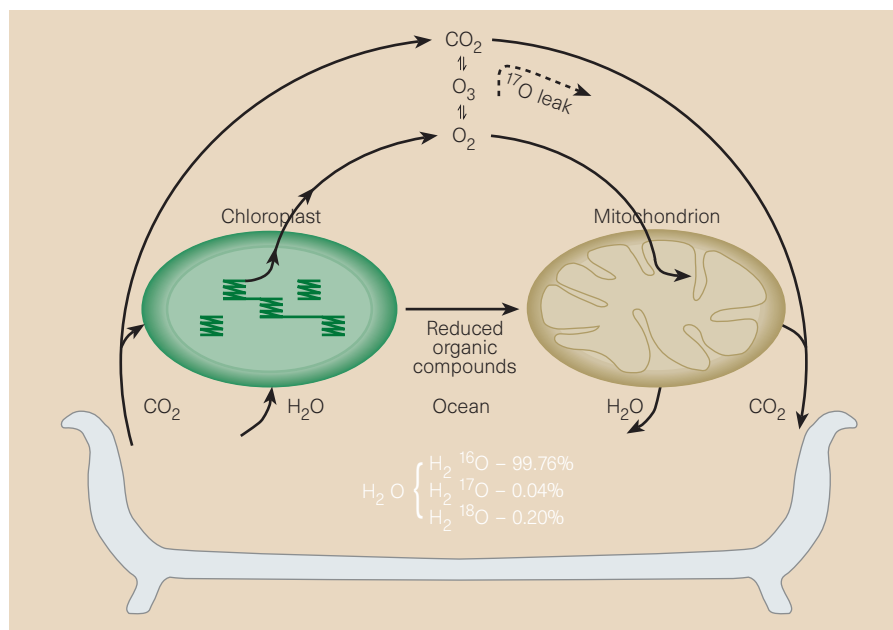


Figure 1 The ^{17}O anomaly. Oxygen in the Earth's atmosphere is continuously recycling; it is produced from water by photosynthesis, in chloroplasts of terrestrial and marine plants, and is consumed by the mitochondria of heterotrophic organisms as they respire reduced organic compounds. Kinetic isotope effects associated with the biological cycle increase the abundance of ^{18}O and ^{17}O in O_2 over that in sea water³. The fractionation of reactions is mass dependent. So, changes in ^{17}O are roughly half those of ^{18}O . Non-biological reactions in the stratosphere have a very different isotope effect, preferentially removing ^{17}O by exchange of oxygen atoms between O_2 , O_3 and CO_2 , enriching the CO_2 of the stratosphere in ^{17}O . On returning to the surface of the planet, CO_2 exchanges oxygen atoms with H_2O , completing this branch of the O_2 cycle. So, the O_2 of the atmosphere contains less ^{17}O than biologically cycled O_2 . Luz *et al.*² measure the size of this anomaly in air from ice cores, and use this to calculate the rate of biological production of O_2 on Earth at different times in the past. Other work from Michael Bender (a co-author of this study) and colleagues³ indicates that the ^{18}O content of air may also contain information on the fractional contribution of marine and terrestrial plants to this productivity.

stratospheric O_2 consumption. The rate of the stratospheric process is independent of the climatic factors that affect biological productivity, and, although this rate is not constant, variation can be estimated from independent measurements. So, Luz *et al.* propose that the size of this anomaly (corrected for variation in the rate of stratospheric O_2 consumption) should provide a proxy for variation in the rate of biological oxygen cycling.

The authors use terrarium experiments to show that biologically recycled O_2 and the free atmosphere have different abundances of ^{17}O and ^{18}O . They develop a methodology, standards and a nomenclature to define the ^{17}O anomaly, then measure the size of this anomaly in today's atmosphere and show that its magnitude is consistent with independent estimates for the rates of biological and stratospheric O_2 consumption. Finally, they present measurements of air from ice-core samples from Summit, Greenland, and calculate that planetary productivity varied from 89 to 97% of the present value during glacial, interglacial and interstadial (temporary ice retreat) stages over the past 82,000 years. These 'proof of concept' measurements will

surely be followed by systematic studies, providing a millennial-scale proxy for global productivity over the time spanned by the ice-core record.

Conservation biology

Top dogs maintain diversity

Bernt-Erik Sæther

Species are becoming extinct at one of the highest rates in the Earth's history¹. So a pressing task in conservation biology is to document the effects of species losses on ecosystems. Such knowledge is necessary to draw public attention to the problem, which in turn is a prerequisite for political action. Unfortunately, in most cases it is almost impossible to determine rigorously the ecological consequences of the loss of a species. On page 563 of this issue, Crooks and Soulé² provide a rare example. It shows how the loss of one 'top predator' — the coyote (Fig. 1) — leads to a higher extinction rate of scrub-breeding birds, through an ecological release of smaller predators.

Large carnivores such as the coyote need large home ranges³. In coastal southern Cali-

fornia, where the intensive urbanization of the past decades has destroyed much of the native sage-bush habitat, the coyote is mainly present only in larger fragments of the remaining natural habitat². Soulé and colleagues have previously suggested⁴ that a decrease in the abundance of a top predator will change the species composition of the carnivore community, leading to 'mesopredator release' — an increase in the number of smaller predators. Larger numbers of native (striped skunk, raccoon, grey fox) or exotic (domestic cat, opossum) mesopredators will then kill more scrub-breeding birds, resulting in higher extinction rates. As expected from this hypothesis, Crooks and Soulé found (using a multiple regression analysis) that mesopredator abundance in a

Joe Berry is at the Carnegie Institution of Washington, Department of Plant Biology, 260 Panama Street, Stanford, California 94305, USA.

e-mail: joeberry@biosphere.stanford.edu

- Bradley, R. S. *Paleoclimatology: Reconstructing Climates of the Quaternary* 2nd edn (Academic, New York, 1999).
- Luz, B., Barkan, E., Bender, M. L., Thieme, M. H. & Boering, K. A. *Nature* **400**, 547–550 (1999).
- Bender, M., Sowers, T. & Labeyrie, L. T. *Glob. Biogeochem. Cycles* **8**, 363–376 (1994).
- Thieme, M. H., Jackson, T. L. & Brenninkmeijer, C. A. M. *Geophys. Res. Lett.* **22**, 225–257 (1995).
- Thieme, M. H. *Science* **283**, 341–345 (1999).
- Alley, R. B. & Bender, M. *Sci. Am.* **278**, 80–85 (1998).
- Bender, M., Sowers, T. & Brook, E. *Proc. Natl Acad. Sci. USA* **94**, 8343–8349 (1997).
- Petit, J. R. *et al. Nature* **399**, 429–436 (1999).



Figure 1 A coyote in southern California, photographed by a remotely triggered camera. Crooks and Soulé² found that bird diversity is higher in fragments of southern Californian sage bush where coyotes are present, as they reduce the number of smaller, bird-killing predators.

habitat fragment decreased significantly as the abundance of coyotes rose, and was highest in small fragments where none were present.

Finding a negative relationship between the abundance of coyote and mesopredators does not, however, prove that there is a direct negative interaction between the two groups of species. The fragmentation process could improve habitat quality for mesopredators, leading to their abundance increasing independently of the loss of coyote. But by analysing the variation in coyote abundance across time within single fragments, Crooks and Soulé were able to provide evidence for a causal relationship. The mesopredators, especially cats, reduced their use of fragments when coyotes were present. Furthermore, coyotes often killed cats. This suggests that mesopredators actively avoid coyotes to reduce the risk of becoming prey.

The changes in the structure of the carnivore community, which were caused by habitat fragmentation, strongly affected the composition of the bird community — as predicted by the mesopredator-release hypothesis. When coyotes became rarer in a fragment, the diversity of bird species decreased, probably because of more efficient predation, especially by cats. The decrease was larger in older fragments. These results are a good illustration of a general principle in conservation biology: it takes a relatively long time to properly assess the impact of human interference on ecosystems.

Crooks and Soulé's study² is an important addition to a small, but exclusive, list that lets us see how changes in the composition of carnivore communities affect species at lower trophic levels. Although based on very few studies, these complicated indirect effects seem to be more of a rule than an exception. For example, when red fox populations in the

boreal forests of northern Europe were greatly reduced by an outbreak of sarcoptic mange during the late 1970s and the 1980s, the number of pine martens shot up, most likely because fewer were killed by foxes⁵. The reduction in the fox population was also associated with an increase in the population density of mountain hares and grouse⁶. Similarly, there is experimental evidence that the species composition of the predator commu-

nity strongly affects the way the prey population fluctuates through time⁷.

Large predators get a mixed press. Many hunters consider them as a severe competitor for ungulates, such as deer. However, if Crooks and Soulé's results represent a general pattern they caution against simply labelling the ecological effects of a species as 'good' or 'bad'. For example, we can only speculate whether the replacement of the red fox as the top predator in many Scandinavian boreal ecosystems by the rapidly expanding grey wolf population will lead to lower fox densities, resulting in more hare and grouse because of reduced predation by foxes. The reintroduction, or unfortunate disappearance, of top predators provides a unique opportunity to study complex trophic interactions. The ecological consequences are very hard to predict, but they are likely to be considerable. □

Bernt-Erik Sæther is in the Department of Zoology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

e-mail: Bernt-Erik.Sather@chembio.ntnu.no

1. May, R. M., Lawton, J. H. & Stork, N. E. in *Extinction Rates* (eds Lawton, J. H. & May, R. M.) 1–24 (Oxford Univ. Press, 1995).
2. Crooks, K. R. & Soulé, M. E. *Nature* **400**, 563–566 (1999).
3. Gittleman, J. L. & Harvey, P. H. *Behav. Ecol. Sociobiol.* **10**, 57–63 (1982).
4. Soulé, M. E. *et al. Conserv. Biol.* **2**, 75–92 (1988).
5. Lindström, E. R., Brainerd, S. M., Hellén, J. O. & Overskaug, K. *Ann. Zool. Fenn.* **2**, 123–130 (1995).
6. Lindström, E. R. *et al. Ecology* **75**, 1042–1049 (1994).
7. May, R. M. *Nature* **398**, 371–372 (1999).

Oceanography

An ultimate limiting nutrient

J. R. Toggweiler

Several elements in the ocean — most notably nitrogen, phosphorus, iron and silicate — are essential for the growth of phytoplankton. Like fertilizers applied to the land, the supply of N, P, Fe and Si in seawater upwelling from below, or in airborne dust deposition from above, determines the rate of production of new organic matter at the surface. It is unlikely, however, that all four elements are equally limiting to growth. Is one clearly more important than the others and, if so, on what timescales? On page 525 of this issue¹ Tyrrell makes a strong case that it is phosphorus, in the form of dissolved phosphate (PO_4^{3-}), that is the ocean's 'ultimate limiting nutrient'.

Of all the nutrients in the ocean, PO_4^{3-} has the longest residence time at around 50,000 years². Residence times of the other nutrients are much shorter — 5,000 years for NO_3^- (nitrate)³, 15,000 years for SiO_2 (silica)⁴ and perhaps 30 years for Fe. Phosphate enters the ocean in rivers, as a product of continental weathering; it is lost from the ocean by being buried in sediments. But like Houdini, PO_4^{3-} is remarkable in its ability to escape burial. Only 1% or so of the upwelled phosphorus

taken up by phytoplankton is trapped in sediments⁵. So, considering its rate of input and output from the system, the marine standing stock of PO_4^{3-} is fairly large.

Much of the recent interest in ocean nutrients has centred on the fast-throughput nutrients NO_3^- and Fe which have relatively low stocks in relation to their delivery to the system. Addition of NO_3^- and Fe has been shown to increase phytoplankton biomass dramatically (see ref. 6 for instance). Tyrrell's work makes us appreciate that these dramatic effects are relatively short lived; on long timescales, it is the PO_4^{3-} content of the ocean that sets the standing stocks of the other nutrients.

Tyrrell set himself the goal of building a simple model to describe the relative abundances of NO_3^- and PO_4^{3-} in the ocean. The model is based on the ecological competition between two types of phytoplankton, those that can fix N_2 from the air ('nitrogen fixers') and those that cannot. The nitrogen fixers have an advantage in growth rate when nitrate is scarce, but are limited by a lower growth rate when it is abundant. Tyrrell assumes that both kinds of phytoplankton

construct their biomass with an N:P ratio of 16:1, the so-called Redfield ratio. His model then reproduces the covariance between NO_3^- and PO_4^{3-} observed in sea water, a simple linear trend with a slope of about 15:1 (see Fig. 1 on page 525). The model produces a slight excess of PO_4^{3-} in surface waters when all available NO_3^- has been consumed. This prediction is supported by observations which show that surface sea water often contains low, but non-zero, concentrations of PO_4^{3-} alongside zero NO_3^- . This slight excess makes nitrogen sufficiently limiting that Tyrrell's nitrogen fixers are at an advantage.

The $\text{NO}_3^-:\text{PO}_4^{3-}$ ratio of sea water is smaller than the N:P ratio in phytoplankton biomass because there is a very large sink for NO_3^- in the ocean's interior due to denitrification, the consumption of NO_3^- and production of N_2 by bacteria in areas where oxygen is not present⁷. When denitrification draws the $\text{NO}_3^-:\text{PO}_4^{3-}$ ratio in the surface ocean down to a level where Tyrrell's nitrogen fixers can compete successfully with the non-nitrogen fixers, the production of new NO_3^- by the nitrogen fixers rises to balance the denitrification sink.

The central point to come out of Tyrrell's model is a prediction that the balance between nitrogen fixation and denitrification is ultimately set by the river input of PO_4^{3-} ; more PO_4^{3-} input leads to more denitrification, which in turn leads to more nitrogen fixation. The nitrogen cycle, being more adaptable than the phosphorus cycle, adjusts to the input and loss of the more limiting nutrient. The fact that the most limiting nutrient has the longest residence time in the ocean implies that the ocean is fairly well buffered against sustained changes in nutrient stocks.

Perhaps the most contentious aspect of

Tyrrell's model is that it makes no provision for an important role for Fe. Nitrogen fixers are known to have a large appetite for Fe, and it is widely thought that the rate of nitrogen fixation is at least locally influenced by its availability from atmospheric dust^{8,9}. In the context of Tyrrell's model, however, the global rate of nitrogen fixation cannot be dependent on the input of an element such as Fe which has a short residence time. This is because a sustained pulse of Fe input leading to increased nitrogen fixation would simply 'mine out' all available PO_4^{3-} from the upper ocean, leading to curtailed phytoplankton production.

From the latest work on Fe cycling, it seems that the residence time of total Fe in the ocean may not be as short as previously believed. Some as-yet-unknown process appears to bind Fe to organic ligands so that it is present in subsurface sea water in much larger quantities than expected from the scavenging kinetics of the naked ferric ion¹⁰. Indeed, it may turn out that organisms alter the cycling of Fe in such a way that in the ocean it is regulated by the river input of PO_4^{3-} , just like NO_3^- . □

J. R. Toggweiler is in NOAA's Geophysical Fluid Dynamics Laboratory, Princeton University, PO Box 308, Princeton, New Jersey 08542, USA. e-mail: jrt@gfdl.gov

1. Tyrrell, T. *Nature* **400**, 525–531 (1999).
2. Van Cappellen, P. & Ingall, E. D. *Paleoceanography* **9**, 677–693 (1994).
3. McElroy, M. B. *Nature* **302**, 328–329 (1983).
4. Treguer, P. et al. *Science* **268**, 375–379 (1995).
5. Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Eldigio, Palisades, New York, 1982).
6. Coale, K. H. et al. *Nature* **383**, 495–501 (1996).
7. Gruber, N. & Sarmiento, J. L. *Glob. Biogeochem. Cycles* **11**, 235–266 (1997).
8. Falkowski, P. *Nature* **387**, 272–275 (1997).
9. Karl, D. et al. *Nature* **388**, 533–538 (1997).
10. Johnson, K. S., Gordon, R. M. & Coale, K. H. *Mar. Chem.* **57**, 137–161 (1997).

almost exclusively of the ^4He isotope, but there is another stable isotope, ^3He , a by-product of the production of tritium in nuclear reactors, which has quite different quantum characteristics. This rare isotope has also been extensively studied, both alone and in the form of ^4He – ^3He mixtures (the two isotopes are partially soluble in each other⁵).

Polturak and his team have studied solid ^4He – ^3He 'alloys' (their term) by using nuclear magnetic resonance (NMR). The basic measurement here is the 'motional narrowing' of the NMR resonance line of ^3He , which allows the speed of diffusion of this isotope to be assessed. Over a narrow temperature range, the helium 'alloy' consists of a mixture of solid plus superfluid in equilibrium, and, because the two phases are of different isotopic composition, the two He isotopes can be regarded as distinct components (at least in terms of the phase rule). Three years ago⁴ Polturak's group made the surprising discovery that the mobility of ^3He in the solid is higher than in the superfluid. No other solid is known in which atoms move faster than in the melt at the same temperature: we have here a remarkable case of superdiffusion, to add to the other 'super' characteristics of helium at low temperatures. It turns out that the solubility of ^3He in ^4He increases hand in hand with the appearance of crystal vacancies — the authors suggest that the ^3He is attracted to vacancies because ^3He has a larger zero-point vibration than ^4He , and so should migrate preferentially into a vacancy. Something similar had been deduced some years ago⁶ from the anomalously fast motion of Sn impurities in solid Pb.

In two sets of experiments^{1,2}, Polturak's group observed slow plastic flow (creep) in single crystals of solid ^4He and of ^4He – ^3He mixtures. Stress was imposed on the solid helium (without any heating) by passing a current through a superconducting metal wire embedded in the crystal in the presence of a magnetic field. The diffusion of vacancies and the counterflow of atoms allowed the wire to move through the crystal (Fig. 1a). A separate coil was used as a sensor to measure the rate of movement of the wire, and velocities as low as 10^{-8} mm s⁻¹ could be measured. For low stress, the creep rate was proportional to the imposed stress, and this is consistent with deformation by the well-established Herring–Nabarro mechanism, also called 'vacancy creep' (Fig. 1a).

The velocity of the wire through a pure ^4He crystal reaches a very high peak due to superdiffusion at the exact temperature at which the h.c.p structure changes to b.c.c. (Fig. 1b)¹. (Various tests prove that the phase transition and the diffusion peak coincide within 1 mK.) Similar maxima, though far less extreme, are seen in b.c.c. metals such as Ti, and have been attributed to the so-called

Materials science

Superdiffusion in solid helium

Robert W. Cahn

Low-temperature physics is a field that can appear sealed off from much of the rest of science, and metallurgy in particular. Its practitioners think in terms of millikelvins rather than, like metallurgists, in thousands of kelvins, and observe strange quantum-linked phenomena such as superfluidity and superconductivity. But when a group of low-temperature physicists — in this case, Emil Polturak and colleagues at the Technion in Israel — focus, exceptionally, on the mechanical behaviour of solid helium their results hold particular interest for metallurgists^{1–3}. Following on from their discovery of unusually rapid diffusion in solid helium⁴, Polturak and co-workers have shown that this 'superdiffusion' occurs at the exact temperature at

which solid helium changes crystal structure¹ and is linked to a reduction in resistance to stress at this temperature². In a computer simulation³, they relate this anomalous behaviour to a new theory of melting. This work covers a whole range of phenomena normally studied by metallurgists, but all within one or two kelvin of absolute zero.

Helium cannot be solidified at ambient pressure, but at pressures exceeding 25 atmospheres and below about 2.5 K it can be turned into a hexagonal-close-packed (h.c.p.) crystal. Over a narrow range of pressures and temperatures near 30 atmospheres and 1.8 K, this structure changes allotropically to a body-centred cubic (b.c.c.) form⁵. Natural helium consists

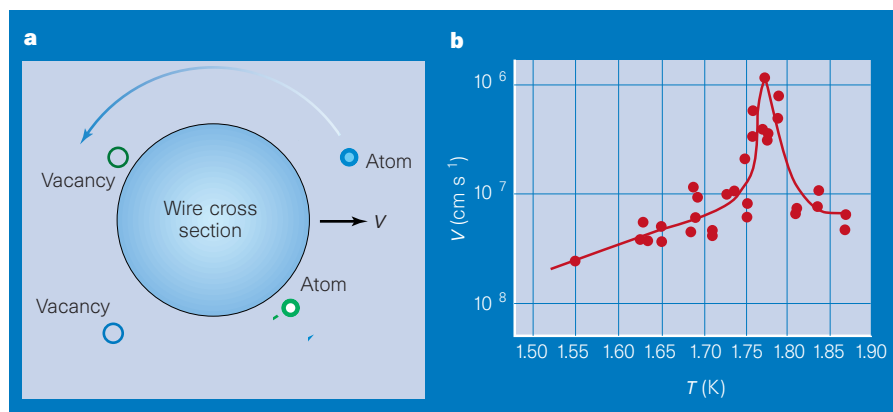


Figure 1 Superdiffusion in solid helium. a, Migration of crystal vacancies, and counterflow of atoms, around a stress-imposing wire in a solid crystal of helium, which allows the wire to move. b, Velocity of the wire as a function of temperature as observed by Polturak *et al.*¹, showing a peak at the temperature where 'superdiffusion' of atoms and vacancies occurs. This peak coincides with the temperature at which helium changes crystal structure, and is much larger than any such effect seen in metals (note that the ordinate scale is logarithmic).

softening of a particular kind of crystal vibration, or mode in the phonon spectrum⁷. Polturak's group thinks the same process causes the helium effect. The extraordinary magnitude of the peak in Fig. 1b is associated with the very low formation energy for crystal point defects in helium (~ 1 meV) compared with b.c.c. metals (~ 1 eV). This low formation energy for He suggests that both the concentration and the mobility of point defects, such as vacancies, might be greatly enhanced near the peak. Computer simulations by Polturak *et al.*³ suggest that the point defects responsible for the very fast diffusion may in fact be a type of interstitial structure in the crystal lattice rather than vacancies.

The acceleration of vacancy creep under a small stress (also called superplasticity by metallurgists) at a phase-transformation temperature is a well-established fact, but in metals the magnitude of such acceleration is nothing like that observed in helium. Polturak and colleagues² have evidence that the presence of ^3He enhances superdiffusion in ^4He by further reducing the point defect formation energy. So in isotopic mixtures, the magnitude of the diffusion peak at the transition temperature is even greater than that shown in Fig. 1b.

The significance of these studies in helium is the unprecedented degree of enhancement of point defect concentration and mobility at a phase transformation. These findings are also relevant to melting theory, in that a very high density of point defects combined with a softened phonon mode can lead to melting, by reducing the shear resistance near the transition to the point where the solid becomes mechanically unstable. In their latest work³, the authors suggest a feedback mechanism in which the point defects in helium soften the phonon (vibrational) spectrum, and this in turn enhances diffusion and creates more point

defects. This feedback mechanism, and a very high density of point defects to begin with, are crucial ingredients in producing a mechanical instability sufficient to generate melting: both are missing in rival theories of melting, of which there have been many over the years. □

Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB3 0DT, UK.
e-mail: rwc12@cam.ac.uk

1. Berent, I. & Polturak, E. *Phys. Rev. Lett.* **81**, 846–849 (1998).
2. Berent, I. & Polturak, E. *J. Low-Temp. Phys.* **112**, 337–354 (1998).
3. Polturak, E. *et al. Physica B* (in the press).
4. Schuster, I. *et al. Europhys. Lett.* **33**, 623–628 (1996).
5. Guyer, R. A. in *Encyclopedia of Physics* 2nd edn (eds Lerner, R. T. & Trigg, G. L.) 486–494 (VCH, New York, 1991).
6. Haskel, D. *et al. Phys. Rev. B* **47**, 14032–14043 (1993).
7. Petry, W. *et al. Phys. Rev. B* **43**, 10933–10947 (1991).

errata In J. Bland-Hawthorn's article "Clues to galaxy formation" (*Nature* **400**, 220–221; 1999) the image in Fig. 1 was not produced by the Virgo Consortium and was not published in ref. 12 as stated. The image should have been attributed to G. Kauffmann, J. M. Colberg, A. Diaferio & S. D. M. White, *Mon. Not. R. Astron. Soc.* **303**, 188–206 (1999), as part of the GIF project (http://www.mpa-garching.mpg.de/~jgc/sim_gif.html).

The source of the cichlid images used in the graphic accompanying Tom Tregenza and Roger K. Butlin's "Speciation without isolation" (*Nature* **400**, 311–312; 1999) should have been acknowledged as a photograph taken by Michael K. Oliver. The photograph of the species concerned, *Dimidiochromis compressiceps*, and others can be viewed at *The Cichlid Fishes of Lake Malawi, Africa* <http://www.connix.com/~mko/>

Daedalus

The insect plane

Aviation engineers look with envy on birds and especially insects. Their flapping wings lift and propel them far more efficiently than the fixed wings of aircraft. One reason is their ability to exploit the subtleties of stalling.

If the angle of attack of a wing is increased, it ultimately stalls, with sudden, disastrous loss of lift. No fixed-wing aircraft dare risk stalling. But an insect with oscillating wings can exploit an intriguing loophole in the laws of aerodynamics. Accelerated at a high angle of attack into the stalling regime, a wing takes a short while to stall. And until it does, it generates enormous lift. By speeding into stall and out again at each flap, an insect wing develops amazingly high average lift.

So Daedalus is inventing a non-steady-state aircraft wing. A conventional wing could never be made to flap, of course. But it might be covered with a flexible elastic fabric, and this could be flapped by a system of rapid repeated inflation and collapse. It might even be made to flap spontaneously in the slipstream, as a flag does in the wind. But the ideal solution is simpler still. Instead of flapping the wing or its surface, Daedalus plans to flap the airflow around it.

Cunningly, he will generate this non-steady airflow from a non-steady propulsive source, a pulse-jet of the type used to power the old V1 missile. Its primitive motor drew in air through a one-way valve, and mixed it with petrol vapour in its combustion chamber. When the chamber was full, a spark ignited the mixture. The valve closed, directing the propulsive blast out through the tail-pipe. The valve then opened and the cycle repeated. So Daedalus's new 'pulse-wing' aerofoil has a leading edge enclosed in cunningly shaped ducting, which acts as a long, thin pulse-jet combustion chamber stretching the length of the wing.

Each time the chamber fires, a sheet of hot gas blasts from the ducting, entraining the airflow round the wing. It speeds it up dramatically, and veers it upwards to put the wind into brief extreme stall, thus creating a sudden pulse of enormous lift. The craft is both propelled and held aloft by repeated pulses. To minimize noise and vibration, Daedalus hopes to drive his pulsed wing in a continuous, distributed manner. Each explosion will spread from the wing root out to its tip, by which time another explosion will be starting at the root.

David Jones

Obituary

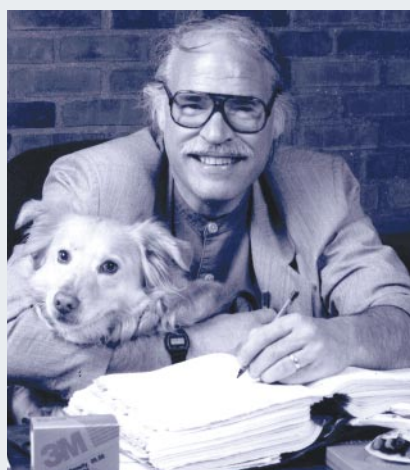
J. John Sepkoski Jr (1948–99)

Palaeontologist who analysed and modelled the history of life

Jack Sepkoski, who died suddenly of heart failure on 1 May, demonstrated more than anyone how the data collected by generations of palaeontologists provide a basis for investigating patterns in the history of life. He matched neither of the stereotyped images of a palaeontologist — the macho explorer and the student of dusty specimens in museum drawers. His field area was, unashamedly, the library.

Sepkoski's work on the history of diversity started in the 1970s at the University of Rochester, New York. He was strongly influenced by the equilibrium model formulated by Robert MacArthur and Edward O. Wilson in studies of island biogeography (that is, the number of species reaches a steady state when originations are balanced by extinctions). Sepkoski's kinetic model, extrapolated back into geological time and considered on a global scale, predicted an exponential rise in faunal diversity after the origin of multicellular animals, followed by a constant equilibrium. Sepkoski showed that the fossil record of marine taxonomic orders from the late Precambrian — around 600 million years (Myr) ago — to the present is remarkably consistent with his model. This provides a simple explanation for the radiation of animal life into major groups during the Cambrian in terms of exponential diversification. But Sepkoski recognized that orders are only a crude proxy for species (the taxonomic hierarchy ranges through phylum, class, order, family and genus, to species).

Sepkoski's analysis of family-level data showed that the history of diversity is more complex. An initial equilibrium, reached by the Late Cambrian (around 515 Myr ago), was followed by a more significant diversification that peaked in the Late Ordovician (about 450 Myr ago). Sepkoski's two-phase kinetic model involved separate evolutionary faunas — one characteristic of the Cambrian, the other of the rest of the Palaeozoic (510–245 Myr ago). The first was dominated by ecological generalists, and the second by more specialized taxonomic groups. This analysis confirmed that a plateau of diversity characterizes the Palaeozoic, and emphasized the profound effect of the extinctions at the end of the Permian (245 Myr ago). Sepkoski later confirmed that a third evolutionary fauna, the modern



fauna, radiated mainly after this end-Permian extinction.

Sepkoski's plot of diversity through time has become the familiar image of the history of life. But as his research hit the limelight in the late 1970s he became something of a *bête noire* to a generation of older traditionalists, who complained that he did not have the rigorous taxonomic training required to ensure the reliability of his data. With characteristic irreverence, Jack caricatured this attitude at one meeting in the mid-1980s by wearing a badge bearing the initials TCCT — Taxon Counter: Casual Theorist! But Sepkoski was no casual theorist; he had the mathematical skills to model and interpret major patterns in the fossil record.

Moving to the University of Chicago in 1978, Sepkoski turned his attention to extinctions, largely in collaboration with David Raup. Major extinctions disrupt equilibria, although the rebound patterns fit kinetic models of diversification. Raup and Sepkoski first identified mass extinctions statistically as peaks that are different from normal or background levels. They also showed that extinction levels have fallen through geological time, implying increased optimization of fitness.

If Sepkoski's models of diversification generated controversy in some quarters, it was nothing compared with the effect of the bombshell that he and Raup dropped in 1984. They announced that there is a significant periodicity in mass extinctions over the last 250 Myr. Their preliminary time-series analysis revealed that the 12 major extinctions occurred at regular intervals of 26 Myr, a discovery that created waves well beyond the palaeontological community. Astronomers proposed Solar-System or Galactic mechanisms, including a hypothetical companion star to the Sun (Nemesis), which precipitated showers of

comets on the Earth every 26 Myr. Other commentators suggested that the periodicity might be an artefact of the way in which fossil ranges are recorded, of the nature of higher taxonomic groups, or of the methods of statistical analysis. But Sepkoski robustly defended the periodicity hypothesis with a much larger database (of genera). He was careful to emphasize, though, that the demonstration of periodicity does not resolve the probable cause of extinctions, or even whether they were gradual or catastrophic events.

Sepkoski's analyses of the fossil record were by no means confined to marine invertebrates. He investigated the dispersal of mammals from the Tertiary period across the Panamanian land bridge (the Great American Interchange, which began around 3 Myr ago). With Conrad Labandeira he demonstrated that high insect diversity is the result of low extinction rather than high origination rates. The radiation of insects was not linked with the evolution of flowering plants, as commonly assumed, but began 100 Myr before their peak.

Sepkoski's main legacy is the patterns that he revealed in the history of life, and his analytical interpretations of diversity dynamics. He published a paper in 1993 entitled "Ten years in the library: new data confirm paleontological patterns" (*Paleobiology* 19, 43–51), which demonstrated that his conclusions on diversification and extinction remained robust despite considerable refinement of the data, including additions and corrections. The fact that these patterns are real is essentially due to the size of the database that Sepkoski painstakingly amassed. But confirmation was in no sense stagnation. Sepkoski continued to investigate issues in the history of life — most recently, rates of speciation, the dynamics of competitive interactions, and the relationship between diversity and ecological importance on a geological timescale.

Jack Sepkoski's research method — long hours holed up with palaeontological monograph and computer — belied his outgoing nature. He was equally at home addressing a major scientific meeting about his latest results. His generosity with both data and ideas, and his sense of fun and enjoyment of company, will be sadly missed. Palaeontology and evolutionary studies have been significantly set back by his untimely death.

Derek E. G. Briggs

Derek E. G. Briggs is in the Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK. e-mail: D.E.G.Briggs@bristol.ac.uk

The tale of the great cuckoo egg

The idea that viruses might cause tumours inadvertently led to the discovery of cellular growth genes that can promote cancer formation — but usually in the absence of viruses. This success rapidly outgrew the original idea.

George Klein

The identification of specific micro-organisms as causative agents of infectious diseases generated much optimism around the turn of the last century. It raised hope that diseases of unknown origin, including cancer, could soon be explained in a similar way. When Peyton Rous showed in 1911 that fowl sarcomas could be transmitted with cell-free filtrates, it was believed that the problem was solved: cancer was a viral disease. But, when similar experiments with mouse and rat tumours failed soon after, it was concluded that tumour viruses occurred only in birds, and the field fell into disrepute.

For the next four decades this opinion prevailed. Potentially important positive findings were 'interpreted away'.

The discovery in the 1920s of the Shope papilloma virus, which causes warts in rabbits, caused little enthusiasm because the tumours were largely benign.

The mouse mammary tumour virus (MMTV), discovered in the 1930s, was called 'milk factor', rather than 'milk virus', to avoid a negative reaction from those — grant-giving bodies included — who had relegated tumour viruses to the cabinet of freaks. Those who understood the viral nature of the 'factor' remained unenthused because analysis showed that MMTV was neither necessary nor sufficient for tumour induction. It merely increased the probability of breast cancer in hosts with both a susceptible genetic background and the appropriate hormonal environment.

The great change in the climate of opinion came in the 1950s when Ludvik Gross discovered the mouse leukaemia virus, and Sarah Stewart and Bernice Eddy identified the polyoma virus. Within a few years the pendulum had swung to the opposite extreme. After decades of failed attempts, viruses that could induce tumours in mammals were now isolated in quick succession. Tumour virology rapidly became a 'most favoured nation' with the grant-giving bodies.

The oncogene concept — that tumour viruses carry 'cancer genes' that can transform some of their target cells into a cancerous or pre-cancerous state — was formulated in the context of this enthusiasm. Ironically, the concept was based on the slow acting or chronic (class II) RNA tumour

By discovering cellular genes that regulate growth and that can contribute to cancer development after illegitimate viral activation, the virologists hatched a great cuckoo egg. These genes, when mutated, can promote cancer formation independent of viruses. Their discovery relegated tumour virology once again to a less prominent place and reaffirmed the sovereignty of cell biology.

virus which, it turned out later, is the only tumour-virus family that does not carry transforming genes. It is the other two families, the directly acting or acute (class I) RNA tumour viruses and the DNA tumour viruses, that carry oncogenes that can transform cells to a cancerous state.

The class I RNA tumour viruses do not cause tumours in nature, nor can they propagate without the intervention of the scientist. They can induce tumours because they have accidentally incorporated from cells genes that regulate growth. Basically, they are highly informative laboratory artefacts. As their name suggests, they carry their genetic information in RNA. After entering a new host cell, the viral enzyme reverse transcriptase copies the viral RNA into DNA (called provirus DNA) which integrates randomly into the host cell's DNA. When the virus starts to reproduce itself, and the proviral DNA is transcribed back into RNA again, some of the new virus particles may also carry additional, cellular sequences from regions adjacent to the virus's random integration site.

Tumour virologists isolated some 20 previously unknown growth regulatory genes from about 40 viral isolates that favour tumour development. These genes originated from corresponding cellular genes which could themselves contribute to spontaneous tumour development, without any viral intervention, after structural or regulatory mutations.

The other type of RNA virus — class II RNA tumour viruses — do not themselves

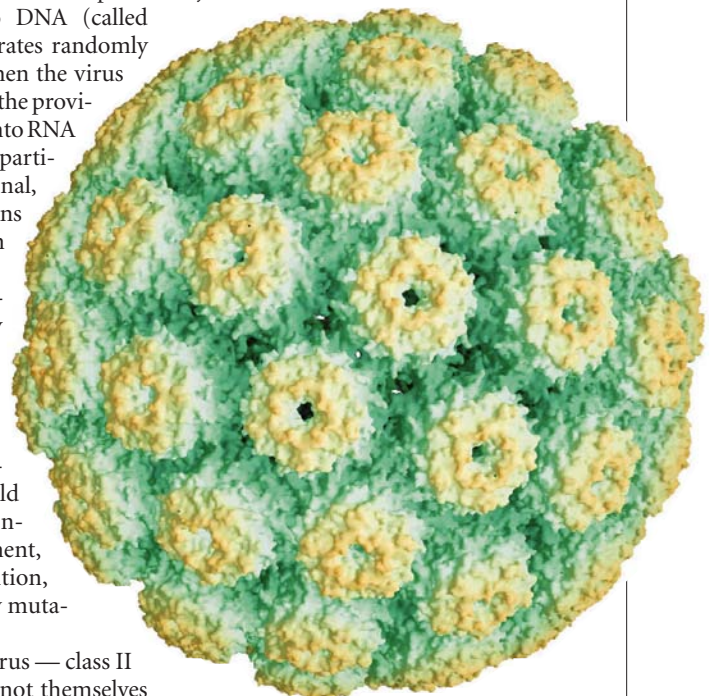
contain oncogenes, but contribute to malignant tumour development relatively infrequently when their proviral DNA happens to integrate in the host DNA near such cellular genes.

By discovering cellular genes that regulate growth and that can contribute to cancer development after illegitimate viral activation, the virologists hatched a great cuckoo egg. These genes, when mutated, can promote cancer formation independent of viruses. Their discovery relegated tumour virology once again to a less prominent place and reaffirmed the sovereignty of cell biology. Cancer is a disease of the cellular DNA.

All oncogenes turned out to be highly conserved household genes that participate in the regulation of the cell cycle. Their potentially tumorigenic forms drive the cell towards division. For overt tumour development, additional genetic changes are required, however. These include the loss of cell-cycle checkpoint controls, inhibition of programmed cell death (apoptosis), and upregulation of blood supply. The oncogene field emerged from erroneous concepts combined with good experimentation. The great cuckoo egg has become a rich source of many unexpected discoveries. □

George Klein is at the Microbiology and Tumorbiology Center, Karolinska Institute, PO Box 280, S-171 77 Stockholm, Sweden.

The SV40 virus does not seem to cause tumours in monkeys, its original hosts, but does when injected into hamsters.



Haem detoxification by an insect

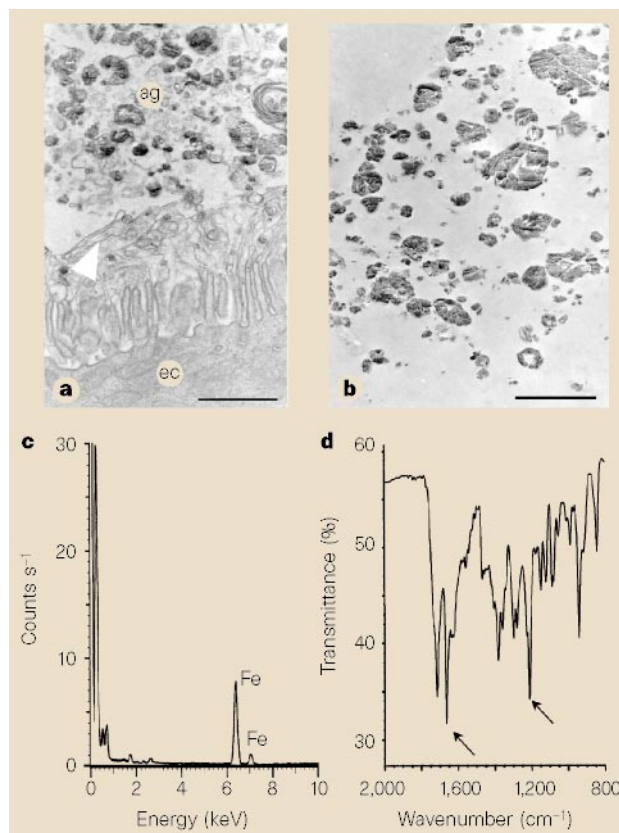
Haem is involved in many biological reactions, including oxygen transport, respiration and photosynthesis. In the free state, however, haem can generate reactive oxygen species that can damage biological molecules. It can also disrupt the phospholipid bilayer of cell membranes¹. In *Plasmodium* parasites, which are the aetiological agents of malaria disease, up to 80% of host-cell haemoglobin is digested², leaving the free haem group to be detoxified in the parasite's food vacuole by polymerizing it into a harmless dark-brown crystalline structure called malaria pigment or haemozoin³. Haem detoxification is also a challenge for blood-sucking insects, which digest several times their own weight of vertebrate blood during a blood meal. Here we show that haem polymerization into haemozoin is not exclusive to *Plasmodium*: it also occurs in the midgut of the blood-sucking insect *Rhodnius prolixus* (Hemiptera), an important vector of *Trypanosoma cruzi*, the causative agent of Chagas' disease.

Transmission electron microscopy (TEM) reveals that the lumen of the *R. prolixus* midgut contains large electron-dense aggregates (Fig. 1a) that are similar in appearance to the haemozoin granules found in *Plasmodium* parasites⁴. These aggregates can be visualized even in the absence of impregnation with electron-dense stains (Fig. 1b), and contain an abundance of iron (Fig. 1c). We conclude that they are derived from digestion of haemoglobin in the insect midgut. After extraction using the same protocol as for malarial haemozoin⁵, this material gives a Fourier transform infrared (FTIR) spectrum very similar to that of malarial haemozoin³, with distinctive peaks at 1,210 and 1,663 cm^{-1} (Fig. 1d). These peaks are not seen in the haemin spectrum and have been attributed to iron-carboxylate bonds in the haemozoin³.

Adducts of haem and acetate anion give a haemozoin-like FTIR spectrum but, like unpolymerized haem, this derivative is soluble in a weakly alkaline solution, whereas haemozoin is virtually insoluble⁶. The material from the *R. prolixus* midgut is completely insoluble at pH 9.1, but it can be solubilized in 0.1 M NaOH as this breaks the iron-carboxylate bonds of haemozoin to produce monomeric haem. These results show that the pigment isolated from the *R. prolixus* midgut is haemozoin. To our knowledge, this is the first time haemozoin has been found in an organism that is not a malaria parasite.

The mechanism of haemozoin formation *in vivo* is controversial. The existence of a haem polymerase enzyme has been proposed on the basis that a crude extract from

Figure 1 Haemozoin in *R. prolixus* midgut. **a**, TEM of a cross-section of *R. prolixus* midgut stained with uranyl acetate and lead citrate, showing electron-dense aggregates in lumen (ag) and epithelial cells (ec). Arrowhead indicates perimicrovillar membranes. **b**, TEM with no staining, showing luminal electron-dense aggregates. **c**, X-ray microanalysis of crystals in the intestinal lumen on nylon grids. **d**, FTIR spectrum of *R. prolixus* pigment from midgut contents 4 days after a blood meal. Arrows indicate peaks characteristic of haemozoin. Midgut contents were obtained by shaking dissected midguts in 0.15 M NaCl. Tissue was discarded and the suspension centrifuged. The insoluble pigment was purified and remaining solids were washed and dried. KBr pellets were prepared from dried samples and spectra were acquired for 32 cycles with a FTIR spectrometer. Scale bars: 0.9 μm (**a**) and 0.6 μm (**b**). Further experimental details are available from the authors.



trophozoites can induce haem polymerization *in vitro*, although no such enzyme has ever been isolated. Also, preformed haemozoin can seed haemozoin formation even in the absence of protein⁷, and it may be driven by phospholipids⁸. In *R. prolixus*, we found that the capacity to induce haem polymerization is associated with a particulate fraction from the midgut lumen that is composed mainly of perimicrovillar membranes, which are extracellular phospholipid-containing membranes that cover the epithelium microvilli and bleb to the intestinal lumen (Fig. 1a).

Haem polymerization in the midgut would be the first line of defence against the effects of releasing haem by haemoglobin digestion. Sequestration of haem into an insoluble form would lead to its elimination in the insect's faeces. Otherwise, in the absence of haemozoin formation, large amounts of haem could cross the midgut wall, potentially resulting in oxidative tissue damage.

The protective effect of haemozoin synthesis in *R. prolixus* is complemented by other mechanisms directed against haem toxicity. *R. prolixus* also has a haem-binding antioxidant protein⁹ and unusually high titres of uric acid (a free-radical scavenger) in its haemolymph¹⁰. Building an array of

defences against haem toxicity therefore seems to be an important trend in the evolution of blood-sucking insects.

Marcus F. Oliveira*, **José R. Silva†**,
Marilvia Dansa-Petretski‡,
Wanderley de Souza‡, **Ulysses Lins§**,
Cláudia M. S. Braga¶, **Hatisaburo Masuda***,
Pedro L. Oliveira*

*Departamento de Bioquímica Médica,
§Instituto de Microbiologia Prof. Paulo de Góes,
and ‡Instituto de Biofísica Carlos Chagas Filho,
Universidade Federal do Rio de Janeiro,
Cidade Universitária,
Rio de Janeiro 21941-590, Brazil
e-mail: pedro@bioqmed.ufjf.br

†Centro de Biociências e Biotecnologia,
Universidade Estadual do Norte Fluminense,
Campos dos Goytacazes,
Rio de Janeiro 28015-620, Brazil

¶Petrobrás/CENPES, Divisão de Química,
Setor de Química Orgânica, Cidade Universitária,
Rio de Janeiro 21949-900, Brazil

- Schmitt, T. H., Frezzatti, W. A. Jr & Schreier, S. *Arch. Biochem. Biophys.* **307**, 96–103 (1993).
- Francis, S. E., Sullivan, D. J. Jr & Goldberg, D. E. *Annu. Rev. Microbiol.* **51**, 97–121 (1997).
- Slater, A. F. G. et al. *Proc. Natl Acad. Sci. USA* **88**, 325–329 (1991).
- Yamada, K. A. & Sherman, I. W. *Exp. Parasitol.* **48**, 61–74 (1979).
- Sullivan, D. J. Jr, Gluzman, I. Y. & Goldberg, D. E. *Science* **271**, 219–221 (1996).
- Pandey, A. V. & Tekwani, B. L. *FEBS Lett.* **393**, 189–192 (1996).

7. Ridley, R. G. *Trends Microbiol.* **4**, 253–254 (1996).
8. Bendrat, K., Berger, B. J. & Cerami, A. *Nature* **378**, 138 (1995).
9. Dansa-Petretski, M., Ribeiro, J. M. C., Atella, G. C., Masuda, H. & Oliveira, P. L. *J. Biol. Chem.* **270**, 10893–10896 (1995).
10. Graça-Souza, A. V., Petretski, J. H., Demasi, M., Bechara, E. J. H. & Oliveira, P. L. *Free Rad. Biol. Med.* **22**, 209–214 (1997).

'Raise the stakes' evolves into a defector

To understand how cooperation can evolve by reciprocal altruism when individuals can make variable investments, Roberts and Sherratt¹ have introduced a new strategy, 'raise the stakes' (RTS), for a continuous version of the iterated 'prisoner's dilemma'. An individual investing I bears a cost I , while the recipient gets a benefit kI . For $k > 1$, this generalizes the standard prisoner's dilemma^{2–5}. Over R alternating encounters^{6,7}, RTS is defined as follows: on the first move, invest a , subsequently raise your investment by $2b$ (or b) if your partner's previous investment bettered (or equalled) your last move, otherwise match your partner's last move. This strategy is denoted by $\sigma = (a, b)$. Roberts and Sherratt¹ reported that the strategy $\sigma = (1, 1)$ performs well in computer simulations against various alternative strategies but did not consider how a population of RTS strategies with different a and b values evolves. We find that selection within RTS populations always acts to lower the values of a and b , hence RTS cooperation is not a robust phenomenon.

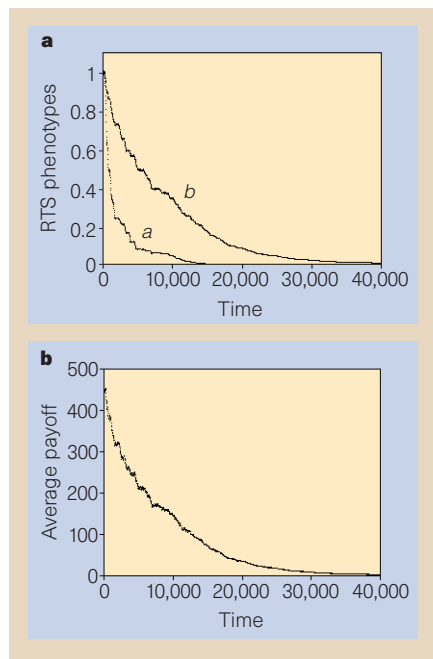


Figure 1 Simulation of the evolution of RTS strategies in the game studied in ref. 1. **a**, Changes in the population mean values of the RTS parameters a and b (starting values, $a = 1$ and $b = 1$). **b**, Change in the mean payoff. In this simulation, $k = 2$, $R = 20$ (the same as in all figures in ref. 1).

Assuming that mutations are small and rare, evolution in a population of RTS strategies can be understood analytically by using adaptive dynamics⁸. If the population consists of individuals using the strategy $\hat{\sigma} = (\hat{a}, \hat{b})$, then the vector field $\xi = \{[\partial S(\sigma, \hat{\sigma})/\partial a]_{\sigma=\hat{\sigma}}, [\partial S(\sigma, \hat{\sigma})/\partial b]_{\sigma=\hat{\sigma}}\}$ determines the direction that optimizes the increase in payoff of a mutant strategy $\sigma = (a, b)$ (ref. 8), where S is the payoff from an iterated interaction. S , and hence ξ , can be calculated analytically and it can be shown that evolution acts to lower the (a, b) parameters of the population. This yields the general prediction that the (a, b) parameters in a population of RTS strategies evolve to zero under natural selection.

This prediction is verified by evolutionary simulations. Consider a population of RTS strategies, with new mutants introduced at a certain rate. In every generation, each strategy plays against all the others and their frequencies in the next generation are calculated using standard game dynamics⁸. Any strategy whose frequency falls below a given threshold is eliminated. A typical simulation (for parameter values used in ref. 1) is shown in Fig. 1. As predicted, the (a, b) parameters evolve to zero. Extensive simulation has confirmed the analytical result for all parameter values studied (including extreme cases, such as $k = 100$, $R = 1,000$).

Thus, in general, RTS evolves under natural selection into an unconditional defector ($a = 0$, $b = 0$). The lack of robustness of RTS arises because, although it is essential from an evolutionary perspective to allow the strategies $\sigma = (a, b)$ to vary continuously (as mutations can, in principle, result in arbitrary changes in a and b), the definition of RTS is discontinuous. From a biological viewpoint, the discontinuous nature of RTS is unrealistic as it is implausible that two strategies that are arbitrarily close would have qualitatively different behaviour.

Although reciprocal altruism with variable investments is an important approach to understanding the evolution of cooperation, our results indicate that new strategies are required to give a satisfactory theoretical account of this process. We have found, both analytically and by simulation, that investment strategies based on an individual's payoff in the previous round (see those used to study mutualism in ref. 9), rather than on the partner's investment, are evolutionarily robust and show how intraspecific cooperation can emerge with variable investments. We believe that these payoff-based strategies represent a more fertile area for future research than RTS strategies.

Timothy Killingback, Michael Doebeli
Zoology Institute, University of Basel,
Rheinsprung 9, 4051 Basel, Switzerland
e-mail: killingback@ubach.unibas.ch

3. Axelrod, R. *The Evolution of Cooperation* (Basic, New York, 1984).
4. Nowak, M. & Sigmund, K. *Nature* **355**, 250–253 (1992).
5. Nowak, M. & Sigmund, K. *Nature* **364**, 56–58 (1993).
6. Nowak, M. & Sigmund, K. *J. Theor. Biol.* **168**, 219–226 (1994).
7. Frean, M. *Proc. R. Soc. Lond. B* **257**, 75–79 (1994).
8. Hofbauer, J. & Sigmund, K. *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press, 1998).
9. Doebeli, M. & Knowlton, N. *Proc. Natl Acad. Sci. USA* **95**, 8676–8680 (1998).

Sherratt and Roberts reply — Killingback and Doebeli argue that our cooperative strategy 'raise the stakes'¹ (RTS) can be continually undermined by selection for less generous strategies. They suggest that the "lack of robustness of RTS" arises from our use of a discontinuous strategy. However, this cannot be the case because the instability they report was in their reformulation of our model in continuous terms. Whether a continuous model is "essential" is debatable. Discontinuous strategies can be more realistic, particularly when resources are not infinitely divisible, hence our notion of a minimal non-zero investment of one unit.

We have also considered the relative success of rare, mutant continuous RTS strategies, but our analyses show that the mean initial investment parameter a will always evolve upwards. Simulations confirm this. Therefore, after trying to replicate their approach, we can find no evidence that even the continuous form of RTS-based cooperation can be eroded in the way they suggest. From this, we cannot exclude the possibility that they have misinterpreted the way RTS operates.

Killingback and Doebeli appear to agree that cooperation can thrive in variable investment systems and that successful strategies would tend to exhibit some initial build up of 'trust'. However, they claim that a strategy that depends on responding to the payoff would be more stable, which we question for two reasons. First, payoff dependency can lead to unnecessary investment in a sucker. Second, in a recent payoff-dependent model², negative payoffs always resulted in the end of cooperation, whereas RTS can rebuild relationships. These sources of instability are highlighted by the need for spatial structuring before payoff-dependent mutualism could evolve². Such assumptions are not required when cost-dependent mutualistic strategies are considered (unpublished data). (Further details are available from T.N.S.)

Thomas N. Sherratt*, Gilbert Roberts†

*Department of Biological Sciences,
University of Durham, South Road,
Durham DH1 3LE, UK

e-mail: t.n.sherratt@durham.ac.uk

†Evolution and Behaviour Research Group,
Department of Psychology,
University of Newcastle upon Tyne,
Newcastle upon Tyne NE1 7RU, UK

1. Roberts, G. & Sherratt, T. N. *Nature* **394**, 175–179 (1998).
2. Doebeli, M. & Knowlton, N. *Proc. Natl Acad. Sci. USA* **95**, 8676–8680 (1998).

Development time and resistance to *Bt* crops

Crop plants genetically engineered to produce insecticidal toxins derived from the bacterium *Bacillus thuringiensis* (*Bt*) are being grown on millions of hectares, but their success will be short-lived if pests adapt to them quickly^{1,2}. The primary strategy for delaying insect resistance to transgenic *Bt* plants is to provide refuges of host plants that do not produce *Bt* toxins. This potentially delays the development of insect resistance to *Bt* crops by providing susceptible insects for mating with resistant insects. But our laboratory results with a worldwide pest of cotton, pink bollworm moths (*Pectinophora gossypiella*)³, contradict an important assumption of the refuge strategy. We find that a resistant strain of larvae on *Bt* cotton takes longer to develop than susceptible larvae on non-*Bt* cotton. This developmental asynchrony favours non-random mating that could reduce the expected benefits of the refuge strategy.

The refuge strategy has two critical assumptions: that inheritance of resistance is recessive, and that random mating occurs between susceptible and resistant insects. If resistance is recessive, hybrid first-generation (F_1) offspring produced by matings between susceptible and resistant adults are killed by eating *Bt* plants. If mating is random, initially rare homozygous resistant adults emerging from *Bt* plants are likely to mate with the more abundant homozygous susceptible adults emerging from non-*Bt* plants, producing hybrid F_1 progeny that cannot survive on *Bt* plants. Mathematical models and limited data from laboratory and greenhouse studies indicate that resistance can be delayed substantially when these assumptions are valid⁴⁻⁷.

Previous work on the feasibility of the refuge strategy has focused on inheritance of resistance, spatial proximity of refuges relative to transgenic crops, and refuge size⁴⁻⁷. To achieve random mating, however, resistant adults from *Bt* plants and susceptible adults from non-*Bt* plants must emerge synchronously^{4,7}. We tested the inheritance of resistance and synchrony for the pink bollworm by measuring survival and developmental rates of a laboratory-selected resistant strain, a susceptible strain, and their hybrid F_1 progeny on *Bt* cotton and non-*Bt* cotton.

Consistent with one of the assumptions of the refuge strategy, we find that pink bollworm resistance to *Bt* cotton is recessive. Survival of the hybrid F_1 progeny (2%) was not higher than survival of the susceptible strain (6%), and both were markedly lower than survival of the resistant strain

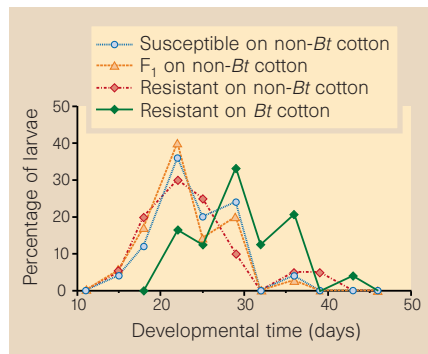


Figure 1 Development of pink bollworm larvae on *Bt* and non-*Bt* cotton plants. Pieces of paper, each with 20–40 eggs, were placed under the bracts of 74 bolls of 16 *Bt* cotton plants (Delta Pine 50B) and 10 non-*Bt* cotton plants (Delta Pine 50) in a greenhouse. After a week, we counted a total of 1,411 entrance holes made by neonates. After two weeks, bolls were caged, and twice a week we counted mature larvae that had exited from bolls. For each strain, survival on *Bt* cotton was estimated by adjusting for mortality on non-*Bt* cotton using Abbott's correction. The resistant strain (APHIS-98R) developed significantly more slowly in bolls of *Bt* cotton (29.8 ± 1.1 days) than the susceptible strain (APHIS-S) did in bolls of non-*Bt* cotton (24.1 ± 0.9) (t -test, $t = 3.89$, d.f. = 47, $P < 0.001$). Developmental time on non-*Bt* cotton did not vary significantly among the resistant strain, the susceptible strain and the F_1 progeny.

(37%) (G -test, $G_{adj} = 24.8$, d.f. = 1, $P < 0.001$). In the only other case in which inheritance of resistance was studied using *Bt* plants, resistance was also recessive⁸. These results differ from the non-recessive resistance to *Bt* toxins in artificial diet seen in a laboratory-selected strain of European corn borer⁹.

Resistant larvae on *Bt* cotton required an average of 5.7 days longer to develop than susceptible larvae on non-*Bt* cotton (Fig. 1). Field data suggest that the median longevity of male pink bollworms is less than a week¹⁰, and laboratory results show that 80% of moths mate within three days of emergence². This developmental asynchrony therefore favours assortative mating among resistant moths from *Bt* plants. In the field, the extent of developmental asynchrony and assortative mating would be affected by variation in toxin expression, weather and overlap between generations.

Assortative mating would generate a disproportionately high number of homozygous resistant insects, accelerating the evolution of resistance. This effect would be diminished if the slower development of resistant larvae increased mortality associated with overwintering or other factors. Computer simulations show that interactions between developmental asynchrony and season length increase uncertainty because they either hasten or slow the evolution of resistance¹¹. There are no reports of resistance to *Bt* crops in the field, but our

results indicate that developmental asynchrony must be considered in efforts to sustain this technology.

Yong-Biao Liu*, **Bruce E. Tabashnik***, **Timothy J. Dennehy***, **Amanda L. Patin***, **Alan C. Bartlett†**

*Department of Entomology, University of Arizona, Tucson, Arizona 85721, USA

e-mail: yliu@ag.arizona.edu

†Western Cotton Research Laboratory, ARS, US Department of Agriculture, Phoenix, Arizona 85040, USA

- Butler, D. & Reichhardt, T. *Nature* **398**, 651–656 (1999).
- Mellon, M. & Rissler, J. *Now or Never, Serious New Plans to Save a Natural Pest Control* (Union of Concerned Scientists, Cambridge, MA, 1998).
- Ingram, W. R. in *Insect Pests of Cotton* (eds Matthews, G. A. & Tunstall, J. P.) 107–148 (CABI, Wallingford, 1994).
- Gould, F. *Annu. Rev. Entomol.* **43**, 701–726 (1998).
- Liu, Y.-B. & Tabashnik, B. E. *Proc. R. Soc. Lond. B* **264**, 605–610 (1997).
- Shelton, A. M., Tang, J. D., Earle, E. D. & Roush, R. T. in *Pest Management – Future Challenges: Proc. 6th Australasian Appl. Entomol. Res. Conf. Vol. 1* (eds Zalucki, M. P., Drew, R. A. I. & White, G. G.) 258–266 (Univ. Queensland, Brisbane, 1998).
- Tabashnik, B. E. *Proc. R. Soc. Lond. B* **255**, 7–12 (1994).
- Metz, T. D., Roush, R. T., Tang, J. D., Shelton, A. M. & Earle, E. D. *Mol. Breed.* **1**, 309–317 (1995).
- Huang, F., Buschman, L. L., Higgins, R. A. & McGaughey, W. H. *Science* **284**, 965–967 (1999).
- Henneberry, T. J. & Keaveny, D. E. III *Suppression of Pink Bollworm by Sterile Moth Releases* (US Department of Agriculture, Springfield, VA, 1985).
- Peck, S. L., Gould, F. & Ellner, S. P. *J. Econ. Entomol.* **92**, 1–16 (1999).

How was the *Sdic* gene fixed?

Nurminsky *et al.*¹ have described a new gene of *Drosophila melanogaster*, termed *Sdic* (for sperm-specific dynein, intermediate chain), which appears to have evolved recently as a result of fusion between duplicated copies of two adjacent genes, *Cdic* and *AnnX*. They consider that the low DNA sequence variation at *Sdic* and *Cdic* is consistent with a recent 'selective sweep' associated with the fixation of *Sdic*. (In a selective sweep, new favourable mutations are incorporated so rapidly that linked alleles can 'hitchhike' and become fixed.) Here we evaluate the evidence for this proposal.

Sdic and *Cdic* are located in the centromere-proximal region of the X chromosome, where the frequency of genetic recombination is greatly reduced². A relation between reduced recombination and reduced occurrence of natural polymorphisms in *Drosophila* is well established³, which Nurminsky *et al.* argue implies frequent selective sweeps¹. But there are other explanations^{4,5}, such as background selection, which can eliminate recurrent deleterious mutations by selection and have the effect of removing variants at linked neutral sites⁴. This model quantitatively fits the data on the relation between recombination rate

and the amount of genetic variation in *D. melanogaster*⁶.

To justify using evidence from DNA sequence variation to show that there has been a selective sweep involving *Sdic* and *Cdic*, it is necessary (if not sufficient) to demonstrate that there is less variation in these genes than would be expected from the background selection model for genes in this region of the X chromosome. This was not done by Nurminsky *et al.*¹, and indeed it is difficult to do. For example, although these genes are in a region where recombination is reduced in frequency compared with the middle of the X chromosome, the relation between the rate of recombination and the location is not known with any accuracy². It is therefore not easy to predict the expected degree of variation under any model, in contrast to the tip of the X chromosome where the gradient of recombination frequency is better known⁶.

The gene *Zw* is located at position D1 on the X chromosome, just distal to *Sdic* and *Cdic*, and has a nucleotide-site diversity (π) of 3.8×10^{-3} (ref. 7); the gene *su(f)* is located at 20E–F and has a π value of 0.5×10^{-3} (ref. 8). The π values for *Sdic* and *Cdic* (0.89×10^{-3} and 0.45×10^{-3} , respectively) are closer to those of *su(f)* than to that of *Zw*, but the large standard deviations of these estimates (0.73×10^{-3} and 0.50×10^{-3} , respectively) mean the true π values may lie between those for *Zw* and *su(f)*. Differences in selective constraints on different genes may also contribute to differences among loci; these can be accounted for by calibrating with respect to interspecific sequence comparisons⁹, which was not done here¹.

The effects of selective sweeps can be detected from departures of frequencies of variants from neutral expectation, as measured by statistics such as Tajima's *D* (ref. 10). The small numbers of variant sites at *Sdic* and *Cdic* mean that such tests lack power in this case. Although Tajima's *D* values are negative for both loci (-0.085 and -0.104 , respectively), indicating an excess of rare variants (as expected after a selective sweep¹⁰), their magnitudes are far below those needed for statistical significance.

We believe, therefore, that the data presented by Nurminsky *et al.* do not provide convincing evidence for a recent selective sweep. This does not imply that *Sdic* has not been fixed by selection. It is possible that the fixation event took place relatively soon after the divergence of *D. melanogaster* and *D. simulans*, and that variability in this region has since returned to that expected for its degree of recombination. Only much better characterization of the levels of genetic diversity and recombination frequencies in this region of the X chromosome can resolve this question.

Brian Charlesworth, Deborah Charlesworth
Institute for Cell, Animal and Population Biology,

University of Edinburgh,
Edinburgh EH9 3JT, UK
e-mail: brian.charlesworth@ed.ac.uk

1. Nurminsky, D. I., Nurminskaya, M. V., De Aguiar, D. & Hartl, D. L. *Nature* **396**, 572–575 (1998).
2. Schalet, A. & Lefevre, G. in *The Genetics and Biology of Drosophila* (eds Ashburner, M. & Novitski, E.) 847–902 (Academic, London, 1976).
3. Begun, D. J. & Aquadro, C. F. *Nature* **356**, 519–520 (1992).
4. Charlesworth, B., Morgan, M. T. & Charlesworth, D. *Genetics* **134**, 1289–1303 (1993).
5. Gillespie, J. H. *Gene* **205**, 291–299 (1997).
6. Charlesworth, B. *Genet. Res.* **68**, 131–150 (1996).
7. Eanes, W. E., Kirchner, M. & Yoon, J. *Proc. Natl Acad. Sci. USA* **90**, 7475–7479 (1993).
8. Langley, C. H., MacDonald, J., Miyashita, N. & Aguadé, M. *Proc. Natl Acad. Sci. USA* **90**, 1800–1803 (1993).
9. Hudson, R. R., Kreitman, M. & Aguadé, M. *Genetics* **116**, 153–159 (1987).
10. Simonsen, K. L., Churchill, G. A. & Aquadro, C. F. *Genetics* **141**, 413–429 (1995).

Nurminsky and Hartl reply — One hallmark of persistent strong background selection is a severe diminution of codon usage bias^{1–3}. An example is the *Drosophila* gene *rolled* (gene 1 in Fig. 1), which is located in the centromeric heterochromatin of chromosome 2, where recombination is severely restricted. Other genes shown in Fig. 1 are located near the base of the X chromosome. Genes 10 and 11 are *AnnX* and *Cdic*, respectively, which flank the *Sdic* gene⁴. As pointed out by Charlesworth and Charlesworth, gene 2, which is *su(f)*, has much less DNA sequence variation than gene 17 (*Zw*). This difference is consistent with the discordant levels of codon usage bias and suggests strong background selection at *su(f)* but not at *Zw*. The degree of codon usage bias shows an extremely sharp increase as the

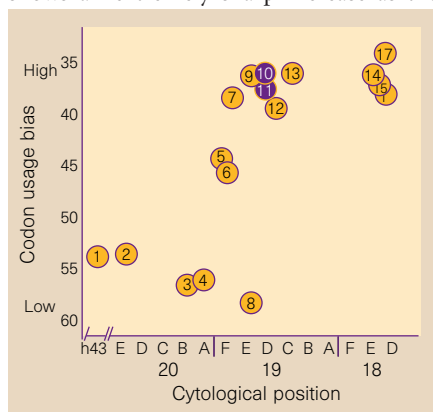


Figure 1 Codon usage bias is scaled according to the effective number of codons⁷. The vertical axis is inverted because a smaller effective number of codons corresponds to a greater bias in codon usage. A similar pattern is seen with other measures of codon bias, such as the χ^2/L statistic⁸ (data not shown). The genes and their accession numbers are: 1, *rl* (M95124); 2, *su(f)* (X62679); 3, *S6kl* (L28945); 4, *fog* (U03717); 5, *sol* (M64084); 6, *slgA* (L07330); 7, *dod* (U35140); 8, *shakB* (U17330); 9, *run* (X56432); 10, *AnnX* (M34069); 11, *Cdic* (AF070699); 12, *Pbprp2* (U05981); 13, *Pp4-19C* (Y14213); 14, *Mer* (U49724); 15, *Cdc42* (U11824); 16, *Bap* (X75910); and 17, *Zw* (M26673, M26674).

gene locations progress outwards from *su(f)*. The transition is near genes 5 (*sol*) and 6 (*slgA*), which are proximal to the *Cdic*–*AnnX* region.

The result is that *AnnX* and *Cdic* are located in a region of codon usage bias similar to that of *Zw*. The cytological region 19DE might therefore support a level of DNA sequence variation of *Sdic* and *Cdic* comparable to that of *Zw*. However, based on the amount of polymorphism observed for *Zw*, the probability of obtaining a value as low or lower than that for *Sdic* and *Cdic* is about 0.043 and 0.008, respectively. These estimates are based on 10,000 simulations using Watterson's formula⁵ for pairwise mismatches in the infinite-alleles model with no recombination, so they should be conservative. There seems to be a statistically significant difference between *Zw* and the other two genes.

The evolution of *Sdic* required an initial duplication and gene fusion accompanied or followed by three deletions, two more insertions/deletions (including one that created a new splice junction), 11 nucleotide substitutions (including reversal of a chain-terminating codon), and a tenfold tandem reiteration of the *Sdic* coding sequence. Although all these changes may have occurred shortly after the divergence of *D. melanogaster* and *D. simulans*, the similarity in degree of codon usage bias of *Cdic* and *Zw*, contrasted with the significant discrepancy in their levels of nucleotide diversity, provides independent evidence in support of our original inference of at least one relatively recent selective sweep. The negative values of Tajima's *D* statistic⁶ for *Sdic* and *Cdic* also support this idea, notwithstanding their lack of statistical significance.

Charlesworth and Charlesworth are correct in pointing out that all the genes in the region 19DE might have limited DNA sequence variation as a result of background selection, despite the higher than predicted level of codon usage bias indicated in Fig. 1. We agree that a much more complete characterization of the levels of genetic diversity and recombination in this region would be informative.

Dmitry I. Nurminsky, Daniel L. Hartl
Department of Organismic and Evolutionary
Biology, Harvard University,
Cambridge, Massachusetts 02138, USA

1. Kliman, R. M. & Hey, J. *Mol. Biol. Evol.* **10**, 1239–1258 (1994).
2. Munte, A., Aguadé, M. & Segarra, C. *Genetics* **147**, 165–175 (1997).
3. Kreitman, M. E. & Antezana, M. in *Evolutionary Genetics from Molecules to Morphology* (eds Singh, R. S. & Krimbas, C.) (Cambridge Univ. Press, New York, in the press).
4. Nurminsky, D. I., Nurminskaya, M. V., De Aguiar, D. & Hartl, D. L. *Nature* **396**, 572–575 (1998).
5. Watterson, G. A. *Theor. Popul. Biol.* **7**, 256–276 (1975).
6. Tajima, F. *Genetics* **123**, 585–595 (1989).
7. Wright, F. *Gene* **87**, 23–29 (1990).
8. Shields, D. C., Sharp, P. M., Higgins, D. G. & Wright, F. *Mol. Biol. Evol.* **5**, 704–716 (1988).

An epidemiology of the invention virus

The Wealth and Poverty of Nations: Why Some Are So Rich And Some Are So Poor

by David Landes

Little Brown: 1998. 650 pp. \$30, £20 (hbk); \$15.95, £10.99 (pbk)

Crispin Tickell

Why are some societies rich and others poor? Why do some seem to advance, and others to stagnate or regress? Are there algorithms, or at least broad formulas, to indicate which paths to follow and which to avoid? Is there a measure of determinism or inevitability in human affairs? In 1776 Adam Smith looked at some of these issues in his *Inquiry into the Nature and Causes of the Wealth of Nations*, and now David Landes, emeritus professor of history and economics at Harvard University, has done the same. He thereby joins the honourable company of world historians, from H. G. Wells and Arnold Toynbee to Paul Kennedy and Jared Diamond, who have jumped the barriers of academic discipline, and tried to see human history as a whole.

His vision is both realistic and disturbing.

Throughout there is a refreshing lack of political correctness. Some parts of the world are more conducive than others to the welfare of the human organism, shaped as it is, like other organisms, by its environment. Climate, resources and geography may not determine human destiny, but they certainly contribute to it. Nature's inequalities (the title of a particularly interesting chapter) defy human attempts to redress them. We have to start any analysis of the past or look into the future on the basis of what actually is, and this rarely fits current ideologies, whether of the free market, central planning or anti-imperialism.

The past 1,000 years, and more particularly the past 500, have seen a process that led to the predominance of Europe, that oddly shaped peninsular at the western end of Asia. The fundamental cause, to which Landes constantly alludes, is the character of European culture. There was never any European monopoly of bright ideas or technical discoveries. Mayan astronomy, Indian mathematics, Islamic science and Chinese practical inventiveness (from the

water clock and gunpowder to the compass, paper and printing) all exceeded European knowledge and technology at the time. The differences arose from the uses to which the Europeans put what they found or took from others.

Exactly why they alone did so is a nice matter for speculation. Among the contributory factors, which varied from place to place and from time to time, must be rich natural resources, diversity of culture within a common culture, dispersal of authority, whether to towns or regions, broad respect for law and private property (at least with regard to other Europeans) and, above all, an individual and entrepreneurial spirit. Together, they made the European ethos distinct from any other.

Landes draws particular attention to the Europeans' use of water wheels as a source of power; the discovery of eye-glasses, which led to spectacles, telescopes and microscopes; the invention of the mechanical clock, enabling time to be measured as never before; the institution of printing presses which led to the first information revolution; and the wide applications of gunpowder which gave the Europeans an unbeatable military advantage and changed the nature of warfare.

Exploitation of these technologies had cascades of political, social, economic and financial consequences. They led to the Renaissance, the expansion of Europe beyond the seas for empire and trade, and the industrial revolution which began in Britain some 250 years ago. The ways in which almost everyone now conducts their lives, whether in wealth or poverty, are a product of these extraordinary events in human history.

The next question is why the virus of invention did not immediately spread elsewhere. Again the answer lies in culture. Some nations were only too willing to learn new ideas and tricks, and to apply them. Journeys of education (or, more precisely, industrial espionage often accompanied by personal inducements) were undertaken widely between the Netherlands, Britain, France, Germany and, later, the United States. Competition between them could not allow otherwise. More often there was strong opposition from established authority, particularly religious bodies. Steel shutters came down to defend society from subversion, as in Spain, later Portugal, Latin America, the Islamic world, Russia and initially Japan. Elsewhere, the virus could not take; the culture was simply too different.

The Chinese had another problem. By the fifteenth century they were drawing in on themselves. For them, any demand for



Have arch, will travel

Arch by Andy Goldsworthy and David Craig (Thames & Hudson, £12.95) is the story of a journey. The 'traveller' is an arch made from red sandstone, and its journey retraces an ancient route by which shepherds drove their sheep to market from Scotland down into northern England. The route is demarcated by sheepfolds,

simple dry-stone enclosures. Many have vanished, to be subsumed by motorways and towns, while others remain. Close to each one, whatever its fate, the arch was erected, stood overnight, and was then dismantled to resume its wanderings to the next fold. Unlike the sheep before it, the arch travelled by truck.

change implied criticism of the Celestial Empire. Their only dealings were with tributaries. When King George III of England wrote to the emperor in 1793 to propose, among other things, a trading relationship, the emperor thanked him for what he described as his homage, but rejected all his propositions, adding "... we have never valued ingenious articles, nor do we have the slightest need for your country's manufactures".

Viruses can usually penetrate sooner or later, but their effects vary enormously. We are witnesses of those changes today. Over time, winners and losers follow each other. The results can be all the more devastating when they are delayed, as in nineteenth-century Japan and twentieth-century China, or when they are disruptive, as in parts of Africa, Southeast Asia, Latin America and the Middle East today. Next come arguments about so-called developed and developing countries, the reasons for wealth in some places and poverty in others, and demands for technology transfer, as if everyone had rights to the benefits of the consumer society.

Landes has fun with the ideologues, particularly certain economists, and enjoys puncturing illusions of all kinds. He does not attempt to lay down doctrines for the future, but instead focuses on the practical realities. These include, on the one hand, convergence of culture in the process loosely called globalization and, on the other, divergence between rich and poor, educated and uneducated, employed and unemployed. However much we may wish otherwise, nature's inequalities are unchanging, and different varieties of culture will continue to resist the homogenization of human society.

This is a marvellous and stimulating book, written with verve, understanding and scholarship. It deserves to be widely read by those in search of the deeper rhythms of human history. I have one broad criticism. As a kind of afterthought, Landes refers to "the serious, progressive, and possibly irremediable damage we are inflicting on the environment". He also remarks, rightly enough, that "other things equal, it is the rich who poison the Earth". To me, this point is fundamental. If more people are to become rich, can the Earth tolerate more poison? What are the likely effects of human population increase; degradation of land; pollution and shortages of water; loss of biodiversity and the natural services we now enjoy for free; climate change (including sea-level rise); and the combination of all these factors? How will the viruses we have released affect not only the human organism but also the ecosystems of which we are a small but immodest part? Perhaps that could be the theme of Landes' next book. □

Crispin Tickell is at Abington Old Barn, Abington, Cirencester GL7 5NU, UK.

From a particular to a global perspective

The Terrestrial Biosphere and Global Change: Implications for Natural and Managed Ecosystems

edited by Brian H. Walker, Will Steffen, Josep Canadell and John S. I. Ingram
Cambridge University Press: 1999. £65, \$110 (hbk); £30, \$49.95 (pbk)

David Schimel

Describing ecology between the 1950s and 1970s in *The Globalization of Ecological Thought* (Ecology Institute, Luhe, Germany), Hal Mooney wrote that "the environment was assumed, although changing in geological time, to vary around a stable mean ... Natural systems were the principal object of study: human-modified ecosystems were more-or-less ignored." He also pointed out that ecology, which began as a global science, became, and to a degree remains, focused on local and small-scale studies. Today's ecologists face problems of global extent resulting from strong anthropogenic trends. Some of the challenges to ecology are intrinsically global, such as understanding the past and future of ecosystems in the carbon cycle. Some problems are becoming global, such as the rapid spread of invasive organisms. Others are local responses to global stresses, of which the leading example is the displacement of the planet's biodiversity by human activities.

Ecologists around the world have responded to the globalization of their science in many ways. The problems involved in studying global ecology are intrinsically different from those of studying, say, the atmosphere. In principle, and to a degree in practice, the atmosphere can be described by basic physical and chemical laws. The essence of ecology, however, is the multi-scaled diversity of the biota: ecosystem types (forests, grasslands, and so on), species diversity within ecosystem types and genetic diversity within species. Few ecological processes can be understood without knowing the particular conditions prevailing at a particular place. Global ecology is built up from a compendium of information about local ecologies and a small but growing number of robust principles that can better knit together local insights into a predictive science.

To achieve a global perspective on a fundamentally heterogeneous world, ecologists, working largely under the auspices of the Global Change and Terrestrial Ecosystems (GCTE) core project of the International Geosphere Biosphere Programme (IGBP), have organized themselves into a number of groups designed to bring site- or region-specific knowledge together to produce global perspectives.

The groups' activities include 'networks' of sites and experimenters who make

comparable measurements under disparate conditions, modellers coming together to compare results and teams who have combed the literature to assemble global databases. In order to understand the global role of ecosystems, measurements must be available everywhere. This has led GCTE to develop an effective strategy for training scientists from less developed nations into world-class scientists who can provide increasingly sophisticated measurements from all corners of the globe.

The results from the first decade of this approach to global ecology are the subject of this book, which also attempts a synthesis of these activities. What emerges is a series of conclusions, rather than an intellectual synthesis, derived from diverse studies. These conclusions provide some important guidelines for thinking about global changes to ecosystems. The book is also a primer of the state of scientific knowledge on biodiversity and ecosystem function.

The focus of GCTE has clearly been on carbon, water and nitrogen fluxes in the plant-soil system at local to global scales. Reviewing the direct effects of carbon dioxide on ecosystems, the authors conclude that there is clear evidence for a significant effect of carbon dioxide on plant water use, photosynthesis, primary productivity and carbon storage. They also state that the effects of carbon dioxide should become saturated and cannot be expected to continue indefinitely. These preliminary conclusions are based on an assessment of a large number of experimental studies, combined with extrapolation using models.

The bulk of terrestrial ecosystems are now 'human dominated', and the crucial role of land use and disturbance in controlling terrestrial biogeochemistry is a major theme of the book. Emissions from change in land use play a major role in the carbon budget, and emissions of other trace gases (especially the nitrogen gases nitrous and nitric oxide) also increase after disturbance. Opposing this for carbon dioxide, regrowth in forests that were harvested 50 to 100 years ago may contribute to the current sink of carbon dioxide in today's ecosystems. This means that systems that were sources of carbon dioxide in the past now appear as sinks. Land use generates a true sink (in the policy sense) only if the recovering systems eventually store more carbon than the pre-disturbance systems. This could happen because of increasing carbon dioxide, nitrogen deposition or climate change, but the data and models do not yet exist to answer these questions adequately.

The GCTE synthesis makes the crucial point that today's terrestrial carbon is precariously close to balance (1.6 gigatonnes of carbon released to the atmosphere from land-use change with an apparent uptake of 1.9). Modest changes in land use or ecosystem physiology (perhaps from additional

warming or drought) could easily change the sign of terrestrial carbon fluxes from a small net sink to a net source. This would further increase difficulties in managing atmospheric carbon dioxide.

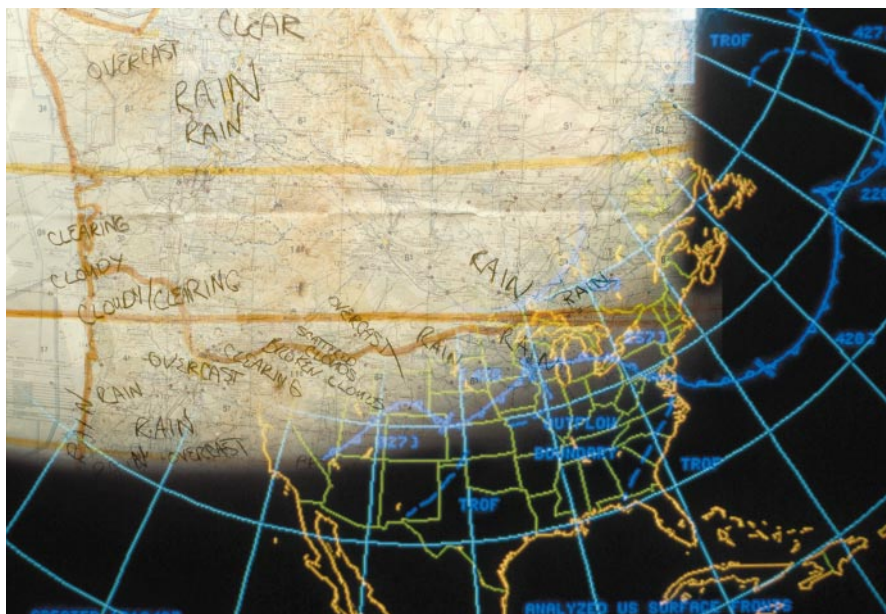
An astonishing number of models are mentioned throughout, but the book leaves the reader confused about the role of models in ecological research. One gets the impression that each group investigating a phenomenon has its own model which possesses idiosyncratic features. Results, then, emerge as consensus or averages over multiple models. The development of many parallel models may be crucial when alternative competing paradigms are being evaluated, or it may, in fact, reflect the historical lack of synthesis in ecology.

In the next round of GCTE synthesis, I hope to see the real points of disagreement and uncertainty emerging more sharply as the basis for the next round of progress. Is the ecological 'credit' system such that scientists feel that, in order to be taken seriously, they must build their own model, even if, in many respects, the resulting models are scientifically interchangeable? GCTE can perform a major role in encouraging scientists to use models as tools in the same way that they use instruments, pervasively but with care for intrinsic limitations. GCTE and the IGBP may also foster the development of common or 'community' models which may then become widely available to scientists in developing and developed nations alike.

This timely review of the part of ecology that deals primarily with material and energy fluxes is also a synthesis of the subdisciplines and is a body of work historically distinct from studies of the distribution and abundance of organisms.

The concluding chapters lay out the likely trajectory of these subdisciplines; the focus on material and energy flow remains. First, it is now recognized that changes to the assemblage of species in a region, including invaders, are a primary control over ecological functioning. This knowledge should strengthen the growing linkage of process-functional ecology to population and community studies. Second, the parallel interests of ecologists and agronomists, rangeland scientists, foresters and soil scientists is leading to a fusion of basic and applied environmental science. Third, the recognition of the role of ecosystem material and energy flows in the climate system is leading to extensive interaction between ecologists and atmospheric scientists.

The science represented within GCTE is a focal point between the deeply biological aspects of ecology (population biology, evolution, ecological genetics) and the other basic and applied Earth system sciences. □
David Schimel is at the Max-Planck-Institut für Biogeochemie, Postfach 10 01 64, D-07701 Jena, Germany.



Mapping the weather: "the problem is to make the mistakes as fast as possible."

A tale of two fronts

Air Apparent: How Meteorologists Learned to Map, Predict, and Dramatize Weather

by Marc Monmonier

University of Chicago Press: 1999. 309 pp.

\$27.50, £21.95

Huw C. Davies

Some people view science as an endeavour to construct maps of reality. On this basis, meteorology would rank among the most prolific (or profligate) of the scientific disciplines. For example, one international weather-forecasting centre alone disseminates more than 150,000 weather charts every day. An individual chart typically portrays the spatial distribution of one or more atmospheric parameter either at the Earth's surface or on some elevated surface. The charts represent the culmination of the daily processing of around 800 megabytes of data from diverse sources — satellites, aircraft, ship buoys, balloons and land-based measurements. These are used to derive an analysis of the current atmospheric state, a detailed multi-day forecast, and an assessment of the reliability of the forecast based on tens of additional but less detailed forecasts.

The challenge is to forecast the future evolution of the inherently unpredictable atmospheric flow — and, moreover, the resulting predictions are subject to daily critical appraisal by both the discipline's practitioners and the general public. No discipline is in a better position to exploit John Archibald Wheeler's dictum that the "problem is to make the mistakes as fast as possible".

Meteorology's evolution has been rapid. In the mid-nineteenth century the fledgling national weather services relied on telegraphed information of the pressure and temperature at a dozen or so predominantly national locations to construct both a hand-drawn surface analysis and an intuitive estimate of weather developments for the subsequent 24 hours.

Acceptance of the discipline after the nineteenth century was promoted by, rather than on, two fronts. The portrayal from the 1920s onwards of cold and warm fronts on weather charts provided the general public with a vivid and readily assimilated depiction of present weather and a hint of future development. And the intriguing possibility that cyclogenesis was attributable to a hydrodynamic instability of an elongated front provided the scientific community with an attractive and acceptable theoretical concept.

So weather maps have been and remain at the centre of meteorology, and they also form the integrating theme of Mark Monmonier's book. The author's avowed intention is to document "the continuing saga of the evolving weather map" and, according to the book's subtitle, to discuss *How Meteorologists Learned to Map, Predict, and Dramatize Weather*. The targeted readership includes both cartographic historians and weather enthusiasts drawn predominantly from the United States. His approach is that of the academic cartographer, and this particular slant both gives the book its charm and defines its weaknesses.

The charm is transparent in the meticulous care with which the author teases out nuggets of cartographic history. These include Edmund Halley's map of the trade winds and monsoons that was published in 1686 (yes, Halley was a meteorologist too),

and the absence of surface-pressure maps attributable to Wilhelm Brandes despite his stipulated plan to achieve this early last century. Likewise, Monmonier gives a detailed description and careful cartographic analysis of weather maps constructed by various acknowledged pantheons of the meteorological community. There is no reproduction of Admiral Fitzroy's highly speculative but remarkably insightful portrayal in the mid-1800s of a family of mid-Atlantic storms. This omission might well be an indication of the author's commendable cartographer's concern for fidelity to the available data.

But the author's underlying cartographic theme is difficult to sustain for two reasons. First, technological developments now provide the forecaster with quasi-continuous data from satellite-based instrumentation and ground-based radar. So cartographic maps become subordinate to film loops, and at one point the author concedes that, for him, "few media are as persuasive as film". In the same vein, his brief consideration of pollutant distribution, ozone depletion and climate change are more indulgent diversions than intellectual excursions. Indeed, his statement that "our best defence against catastrophic climate change rests on a four-fold strategy of monitoring, modelling, archiving, and visualization" should be treated with reservation.

Second, and far more trenchantly, meteorological cartography should be motivated and underpinned by physical considerations. The task is to select the most insightful atmospheric fields and to display them with clarity in two, three or four dimensions. The book has little to say on the relative value of difference atmospheric fields, while the narrative is pedestrian and sparse in its discussion of weather prediction.

Contrast this with the study of the young Halley mentioned above. On the one hand, scholars rank Halley's wind chart among the most important in the history of cartography. On the other, his objective is evident in the title of the accompanying paper — "An historical account of the trade-winds and monsoons observable in the seas between and near the tropics with an attempt to assign the physical cause of the said winds".

A potential buyer will have to assess whether the book's charm outweighs its shortcomings. □

Huw C. Davies is at the Institute for Atmospheric Science ETH, Hönggerberg HPP, CH-8093 Zürich, Switzerland.

More on the atmosphere

The Stratosphere: Phenomena, History, and Relevance

Karin G. Labitzke & Harry Van Loon
Springer, £37.50, \$62

Science in culture

Lucid looking

David Hockney's drawings using the camera lucida
Martin Kemp

What are we to think when one of the greatest draughtsmen of our age resorts to a drawing device patented in 1806 and primarily intended to provide assistance to 'amateurs'? The artist in question is the renowned British painter, maker of photo-works, stage designer and master of the pencil, David Hockney. The device is the camera lucida invented by William Hyde Wollaston, physician turned chemist and optician, as a way of overcoming his inability to depict admired scenery.

The camera lucida stood in a long line of descent from various optical drawing devices, ranging from mechanical perspectographs to lens-based camera obscuras. Indeed, its name played on the fact that it did not need the 'dark chamber' of the camera obscura and could be used in any light conditions. The convenient and readily portable camera lucida used a four-sided prism, two faces of which are set at 135°, to send a twice-reversed image to the eye via two internal reflections. Wollaston explained how a draughtsman could simultaneously see both the object to be portrayed and a horizontal drawing surface, providing the eye is "so placed that only a part of its pupil may be intercepted by the edge of the prism". This difficult trick was facilitated by a hinged eye-hole, while supplementary lenses helped with the problem that the eye was required to focus on both the object and the image plane.

Used by some professional draftsmen, including the sculptor Sir Francis Chantrey, as well as by countless hopeful novices, the camera lucida largely passed out of service with the spread of photography. Wollaston's prism was a demanding device to use, requiring considerable ocular control. Even then, an unskilful artist would achieve only "lamentable results" — to quote William Henry Fox Talbot, whose failure to produce decent landscape drawings with the camera proved to be a powerful goad to his invention of calotype photography. By contrast, the astronomer Sir John Herschel, adept at using his eyes with telescopes and well equipped with draughtsman's skills, achieved beguiling results.

Hockney, who has consistently been concerned with issues of seeing, representation, perspective, space and the camera, has recently



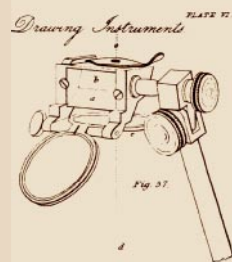
Hockney's camera lucida drawing of Barry Humphries (courtesy of Annely Juda Fine Art), on exhibition at 23 Dering Street, London, until 18 September.

seized on Wollaston's invention to undertake a series of drawn portraits. Unlike an inexperienced artist, who is likely to attempt a laboured tracing of contours, Hockney uses the instrument as a sighting device, briskly demarcating key points of the features, such as the corners of the eyes and line of the mouth. The advantage is that such key registers of expression can be rapidly established before the sitter's expression freezes or sags. Removing the device, he then directly delineates the shades and highlights through an intense process of observation and depiction, in which his gaze incessantly 'tick-tocks' from face to paper at intervals of no more than two seconds. Being portrayed by Hockney is, as I can testify, to be machine-gunned by an ocular marksman of the first order.

The results share something of the quality of photographs, while not looking like photographic images. They remain, recognizably, drawings in Hockney's style, yet they are discernibly different from the entirely 'free-hand' portrait drawings from earlier in his career. They exhibit a combination of fresh immediacy (the short exposure) and unrelenting intensity (the long exposure) that entirely validates the artist's surprising adoption of the camera lucida.

The results also feed back into the history of the device. Hockney took advantage of his recent visit to London from his home in California to scrutinize the portrait drawings by the French Neo-classical artist Jean-Auguste-Dominique Ingres, then on exhibition at the National Gallery. His intuition is that Ingres used a camera lucida to facilitate the making of drawn likenesses of visitors to Rome in the second decade of the nineteenth century. It is a happy intuition that, in my capacity as a historian, I hope to explore further. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.



Optical pieces of a camera lucida, from Cornelius Varley's *A Treatise on Optical Drawing Instruments*, 1845.

The relative influences of nitrogen and phosphorus on oceanic primary production

Toby Tyrrell

School of Ocean and Earth Science, Southampton Oceanography Centre, University of Southampton, European Way, Southampton SO14 3ZH, UK

A simple model has the potential to resolve the long-running debate amongst oceanographers over whether nitrogen or phosphorus exerts overall control on oceanic primary production. A representation of the competition between nitrogen-fixing and other phytoplankton is inserted into a two-box global model of the oceanic nitrogen and phosphorus cycles. Homeostatic regulation of both nitrate and phosphate concentrations results, with surface waters more deficient in nitrate than phosphate in the steady state, but with external phosphate inputs controlling longer-term primary production in the global ocean.

There are two views of how primary production in most of the world's oceans is controlled: the 'geochemists' view (phosphorus regulation) and the 'biologists' view (nitrogen regulation).

When moving down from the surface to the deep ocean, nitrate and phosphate concentrations both increase, and they do so proportionately in the ratio $[\text{NO}_3^-] : [\text{PO}_4^{3-}] \approx 15 : 1$ (Fig. 1). Redfield¹ first realized that this stoichiometric increase is due to the equal rates of release of N and P from sinking particles of decomposing organic matter, which also contain N and P in a ratio of (15–16):1. But he also noted that upon biological depletion of nutrients:

"no great excess of one element is left when the supply of the other has been exhausted. .. That two components of such great importance in the synthesis of living matter are so exactly balanced in the marine environment is a unique fact and one which calls for some explanation, if it is not to be regarded as a mere coincidence. .."¹

Put another way, why does the best-fit line to the data in Fig. 1 intersect the axes close to the origin, rather than a long way along either axis? The activities of nitrogen-fixers may be responsible^{2,3}. When nitrate is scarce relative to phosphate (low $[\text{NO}_3^-] : [\text{PO}_4^{3-}]$) then those organisms which can obtain their nitrogen from the super-available, atmosphere-derived dinitrogen (N_2), will become more numerous⁴. As this N_2 -fuelled algal growth is eaten and decomposed, its nitrogen returns to dissolved inorganic ammonium and nitrate in the water, increasing $[\text{NO}_3^-] : [\text{PO}_4^{3-}]$ in a negative feedback. However, there is no atmospheric reservoir of phosphorus and so there is no alternative source once phosphate runs out. According to this view (the 'geochemists' view), nitrate gets topped up when scarce relative to phosphate, nitrate concentrations therefore slavishly follow phosphate concentrations, and so it is phosphate dynamics that control ocean primary production and fertility. This scheme has been experimentally demonstrated in lakes^{5–8}, but not oceans.

However, large data sets, such as GEOSECS⁹ and the World Ocean Atlas¹⁰, show that in reality nitrate usually runs out slightly before phosphate when nutrients become depleted in surface waters (Fig. 1); that is, surface $[\text{NO}_3^-] : [\text{PO}_4^{3-}]$ is less than 15:1. Nutrient-depleted (oligotrophic) waters usually contain a small residue of phosphate while nitrate is undetectable. Biologists have also found that nitrate additions to samples of nutrient-poor surface ocean water typically stimulate phytoplankton growth whereas phosphate additions do not^{11,12}. Because nitrate is found in practice to be the most limiting nutrient in surface waters, this has led to the conclusion that nitrate rather than phosphate is the 'master limiting

nutrient' (the 'biologist's view'), and that it is the dynamics of the nitrogen cycle which are important for controlling phytoplankton productivity^{12–14}.

This controversy¹⁵ has been an important issue in biological oceanography during the past 25 years, dating from ref. 11. Here I describe a simple but quantitatively reasonable model that allows the reconciliation of these two views.

Proximate and ultimate limiting nutrients

The distinction between the following two definitions is fundamental to this Article:

(1) The 'proximate limiting nutrient' (PLN) represents the local limiting nutrient according to Liebig's law (growth rate is determined only by the availability of the most limiting substrate). Enrichment of a bottle of surface water with just the PLN will result in an enhancement of primary production within hours or days.

(2) In contrast, the 'ultimate limiting nutrient' (ULN) represents the nutrient whose supply rate forces total system productivity over long timescales. If the river input of the ULN changes then total system primary productivity will also eventually change, perhaps over thousands of years for the case of the global ocean.

The PLN relates to the definition⁷ of level I (culture) and level II (community culture) experiments, assuming that the growth medium is *in situ* sea water. The ULN relates to the results of level IV (rivers, lakes, bays, oceans) experiments.

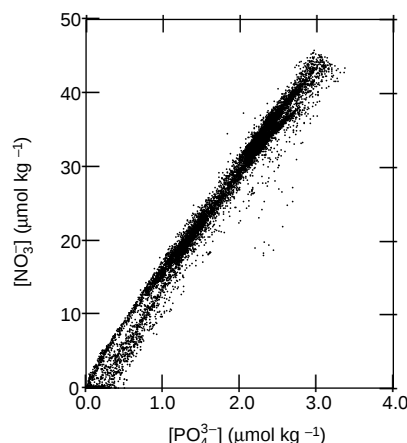


Figure 1 $[\text{NO}_3^-]$ versus $[\text{PO}_4^{3-}]$ scatter plot from GEOSECS global data set⁹.

My model shows that the PLN and the ULN need not be the same: the ocean's PLN is reactive nitrogen while its ULN is simultaneously phosphate. The outcome of this debate is of importance to historical controls on climate (should we be interested in proxies of N-cycling or of P-cycling as controls on past partial pressures of CO_2 in the atmosphere, p_{CO_2} ?), and to continuing attempts at eutrophication control (should N or P be removed from wastewater inputs to sensitive waters?).

Model description

Figure 2 shows how the two element cycles are represented in the model. It is a standard one-dimensional, two-box model of the global ocean¹⁶, with the top layer representing the surface ocean down to the limit of the deepest wind-induced mixing during the year (the annual thermocline) and the bottom layer representing the deep ocean.

Marine phosphorus cycle^{17,18}. Phosphorus is almost totally absent from the atmosphere, and the only significant input of phosphorus to the oceans comes via river water. The most significant output of phosphorus from the oceans is in organic debris sinking to the ocean floor and becoming incorporated into sedimentary rocks. The most important processes responsible for the internal distribution of phosphate in the oceans, are as follows: (1) the assimilation of dissolved phosphate into phytoplankton biomass; (2) the subsequent release (through grazing, cell lysis, bacterial degradation, and so on) of most of this phosphorus back into dissolved nutrients in the surface ocean; (3) the release of a further portion of the organic phosphorus back into deeper waters after organic material has sunk out of the surface oceans; and (4) slow mixing between the surface and deep oceans, partly due to the physical movement of large masses of water between the surface and deep layers (upwelling and downwelling), and partly due to diffusion.

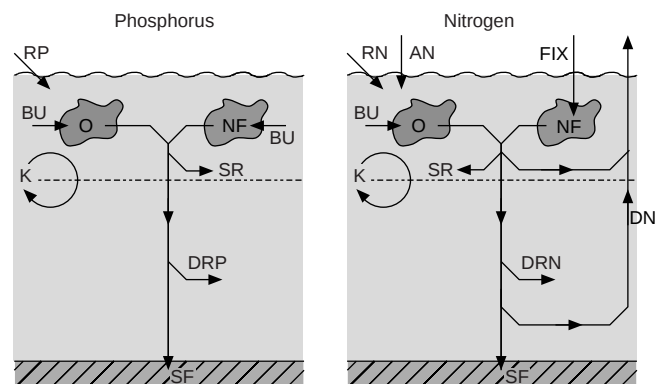


Figure 2 Structure of model. RP and RN, river inputs; AN, atmospheric nitrogen input; BU, biological uptake; FIX, nitrogen fixation; SR, surface regeneration of nutrients; DRP, DRN, deep regeneration of P and N respectively; K, mixing; SF, sedimenting fraction; DN, denitrification; NF, nitrogen-fixing algae; O, other algae.

Marine nitrogen cycle^{18,19}. This is more complex than the phosphorus cycle. Nitrogen exists in several dissolved forms in the ocean, including nitrate (NO_3^- , typically present at $0\text{--}40\ \mu\text{mol N kg}^{-1}$), nitrite (NO_2^- , $0\text{--}1\ \mu\text{mol N kg}^{-1}$), ammonium (NH_4^+ , $0\text{--}1\ \mu\text{mol N kg}^{-1}$), dinitrogen (N_2 , $\sim 1,000\ \mu\text{mol N kg}^{-1}$), and dissolved organic nitrogen (DON) compounds not directly accessible by phytoplankton. 'Reactive nitrogen' is the only form of nitrogen explicitly modelled here, and refers to the sum of the forms of nitrogen that are easily taken up by phytoplankton: NO_3^- , NO_2^- and NH_4^+ . N_2 concentration (requiring greater energy expenditure to access, due to the strong triple bond in N_2) is not explicitly included in the model because it is always present in excess in the water. Another difference from the marine phosphorus cycle is the significant loss of reactive nitrogen due to denitrification, which transfers reactive nitrogen back to N_2 . For simplicity, other N and P processes of lesser importance are omitted from the model, although considered in sensitivity analyses.

Two types of phytoplankton. Marine photosynthesizers which can fix N_2 , such as many cyanobacteria, have an advantage when reactive nitrogen is scarce (if PO_4^{3-} is also available) since they are not reliant on that limited supply. However, they are at a disadvantage compared to other phytoplankton when reactive nitrogen is more abundant, since N_2 -fixation requires more energy (if it were otherwise then all phytoplankton would make use of the large N_2 reservoir and ignore the other forms of nitrogen).

To represent the differences between nitrogen-fixing and non-nitrogen-fixing ('other') phytoplankton in the model, some elements from phytoplankton ecological modelling²⁰ are included. In line with normal practice, nutrient limitation of growth rate is calculated according to the Michaelis–Menten formula (Fig. 3; equations (1) and (2) in Box 1). For the other phytoplankton, the combined limitation due to N and P together is calculated according to Liebig's law.

To encode for the greater cost of nitrogen fixation, the nitrogen-fixing algae are given a slightly lower maximum growth rate. The greater requirements of such algae for iron and molybdenum⁸ are not represented in this model, despite some observational evidence of Fe effects on N_2 -fixer occurrence¹⁴. Both types of algae are subjected to a mortality term, representing losses due to grazing (consumption by zooplankton), sinking and mixing of cells out of the photic zone.

Competing N_2 -fixers and other phytoplankton have been previously modelled²¹ in a one-box estuarine system, and some similar results were produced.

Model equations

There are six variables in my model, representing surface and deep $[\text{NO}_3^-]$ and $[\text{PO}_4^{3-}]$, and two types of phytoplankton in surface water only, and these are represented by the equations shown in Box 1.

Equations (1) and (2) together give the nitrogen-fixers a growth advantage when both reactive nitrogen is in short supply and

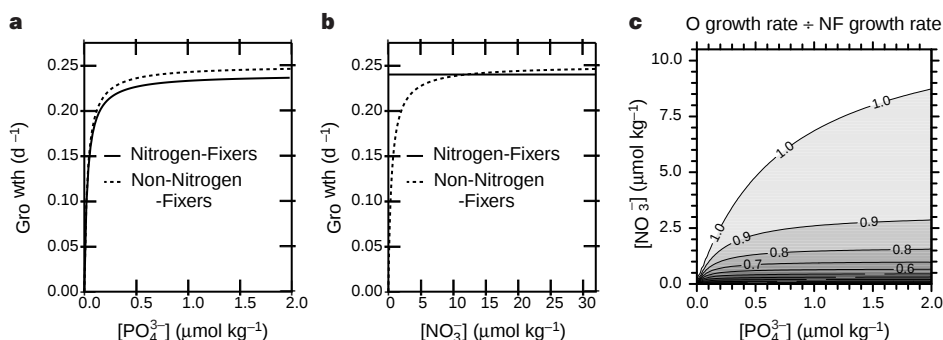


Figure 3 Nutrient limitation of the two phytoplankton groups in the model. **a**, Limitation by phosphate, when nitrate is present in excess (N_2 -fixers have the lower maximum growth rate); **b**, limitation by nitrate, when phosphate is present in

excess (the N_2 -fixers are not limited by nitrate); **c**, ratio of growth rates (indicating competitive dominance) at different $[\text{NO}_3^-]$ and $[\text{PO}_4^{3-}]$. N_2 -fixers will win out against other algae when there is some phosphate present and nitrate is in short supply.

phosphorus is relatively abundant; but these equations put the other phytoplankton at an advantage otherwise (Fig. 3). Growth rate differences translate to abundance differences (out-competition of one phytoplankton group by the other) because both phytoplankton are given the same mortality rate, and because of restricted phosphate availability which sets a limit to the total biomass that can be supported. The calculation of phytoplankton growth rates omits the effects of light, temperature, silicate and micronutrient (for example, iron) limitations on growth rate, all of which are important over some parts of the surface ocean at some times. To compensate, the phytoplankton are given lower maximum growth rates than laboratory and field experiments indicate for optimal

growth conditions.

Considerations of global mass balance for phytoplankton suggest that the average birth/death rate for phytoplankton equals $\sim 20\%$ of the population per day (total (net) primary production, TPP, of 50 Gt C yr^{-1} divided by global biomass of 0.6 Gt C ; Box 2), and this mortality value was used here in preference to larger values suggested by Banse²².

Equations (3)–(6) in Box 1 simulate the processes of riverine and atmospheric inputs to the ocean, assimilation of nutrients into organic matter (new phytoplankton growth), remineralization of that organic material in surface and deep waters, sedimentation of a small portion of the organic material, mixing of nutrients between

Box 1 Model parameter values and equations

Parameter values*

Symbol	Description	Units	Model value	Literature values
SD	Depth of surface layer	m	500	
DD	Depth of deep layer	m	3,230	
K	Mixing coefficient	m yr^{-1}	3.0	3.0 (ref. 2)
SR	Fraction of TPP regenerated above 500 m†	%	95	92 (ref. 29), 97 (ref. 30)
SF	Fraction of TPP permanently incorporated into sediments	%	0.2	0.1–0.2 (ref. 18), 0.16 (ref. 31)
DR	Fraction of TPP regenerated below 500 m (= 100–SR–SF)	%	4.8	
DN	Fraction of TPP of N (= $8,100 \text{ Tg N yr}^{-1}$) that is regenerated via denitrification to N_2	%	1.5	$25\text{--}180 \text{ Tg N yr}^{-1} \approx 0.3\text{--}2.2\%$ (ref. 19), $20\text{--}170 \text{ Tg N yr}^{-1} \approx 0.2\text{--}2.1\%$ (ref. 18), $100\text{--}200 \text{ Tg N yr}^{-1} \approx 1.2\text{--}2.5\%$ (ref. 12), $250 \text{ Tg N yr}^{-1} \approx 3.1\%$ (ref. 13), 0.09×10^{-3} (ref. 32), 0.19×10^{-3} (ref. 18), 0.15×10^{-3} (ref. 17), 0.21×10^{-3} (ref. 33)
RP	River input of P§ (total dissolved)	$\text{mol P m}^{-2} \text{ yr}^{-1}$	0.2×10^{-3}	2.9×10^{-3} (ref. 32), $(2.6\text{--}8.4) \times 10^{-3}$ (ref. 18), $(2.8\text{--}7.9) \times 10^{-3}$ (ref. 19), 3.7×10^{-3} (ref. 33)
RN	River input of N§ (total dissolved)	$\text{mol N m}^{-2} \text{ yr}^{-1}$	6.0×10^{-3}	9.9×10^{-3} (ref. 31), $(4.5\text{--}11.0) \times 10^{-3}$ (ref. 19), $(4.1\text{--}11.0) \times 10^{-3}$ (ref. 34), $(5.9) \times 10^{-3}$ (ref. 35), $(3.4) \times 10^{-3}$ (ref. 33), $(14.1\text{--}59.9) \times 10^{-3}$ – $(19.1\text{--}41.5) \times 10^{-3}$ (ref. 18)
AN	Atmospheric input of N	$\text{mol N m}^{-2} \text{ yr}^{-1}$	7.5×10^{-3}	$16\text{--}17$ (ref. 36), 17 (ref. 37)
R_{ORG}	N:P ratio in phytoplankton biomass	mol/mol	16	
$\mu_{\text{NF}}^{\text{max}}$	Maximum growth rate of N_2 -fixing phytoplankton	yr^{-1}	$0.24 \text{ d}^{-1} = 87.6 \text{ yr}^{-1}$	$0.1\text{--}4.1 \text{ d}^{-1} = 36\text{--}1,500 \text{ yr}^{-1}$ (ref. 38)
$\mu_{\text{O}}^{\text{max}}$	Maximum growth rate of other phytoplankton	yr^{-1}	$0.25 \text{ d}^{-1} = 91.25 \text{ yr}^{-1}$	0.03×10^{-3} (ref. 39), 0.05×10^{-3} (ref. 40)
P_{H}	M–M half-saturation constant for growth vs. $[\text{PO}_4^{3-}]$ ‡	mol P m^{-3}	0.03×10^{-3}	(both <i>Skeletonema costatum</i> only)
N_{H}	M–M half-saturation constant for growth vs. $[\text{NO}_3^-]$ ‡	mol N m^{-3}	0.5×10^{-3}	0.5×10^{-3} (ref. 41), $(0.1\text{--}4.2) \times 10^{-3}$ (ref. 42)
M	Mortality (mainly grazing)	yr^{-1}	$0.20 \text{ d}^{-1} = 73 \text{ yr}^{-1}$	$0.25\text{--}1.2 \text{ d}^{-1} = 91\text{--}440 \text{ yr}^{-1}$ (ref. 22)

* Where literature values were given in different units from the ones stated then the following assumptions were used to convert them: (1) all organic matter is in the Redfield ratio, that is, C:N:P = 106:16:1; (2) molar weights of C, N and P are 12, 14 and 31 g, respectively; (3) area and volume of the global = $361 \times 10^{12} \text{ m}^2$ and $135 \times 10^{16} \text{ m}^3$; (4) $1 \text{ mol kg}^{-1} = 10^{-3} \text{ mol m}^{-3}$; (5) phytoplankton specific growth rate = $0.693 \times$ phytoplankton doubling rate; (6) TPP = $8,109 \text{ Tg N yr}^{-1}$.

† The fraction of TPP being regenerated in the surface layer depends critically on the assumed depth of that surface layer; only literature estimates for $\sim 500 \text{ m}$ surface layers are included here.

§ Estimates of river flux depend on whether just dissolved forms (such as DIN + DON) are included, or whether the particulate load is also included. The literature estimates quoted are just for dissolved river inputs, although the parameter used is slightly increased to take into account the possible availability of part of the particulate load.

‡ The M–M (Michaelis–Menten) half-saturation constant for a nutrient is the concentration of that nutrient at which either (1) the growth rate, or (2) the nutrient uptake rate, for the species consuming it is half of its maximum value. The literature values given in the table are for nutrient uptake whereas the model values are for growth because cell quotas are not modelled.

Equations

$$d(\text{NF})/dt = +\mu_{\text{NF}}^{\text{max}} \frac{P_{\text{S}}}{P_{\text{S}} + P_{\text{H}}} \cdot \text{NF} - M \cdot \text{NF} \quad (1)$$

$$d(\text{O})/dt = +\mu_{\text{O}}^{\text{max}} \cdot \text{minimum} \left\{ \frac{P_{\text{S}}}{P_{\text{S}} + P_{\text{H}}}, \frac{N_{\text{S}}}{N_{\text{S}} + N_{\text{H}}} \right\} \cdot \text{O} - M \cdot \text{O} \quad (2)$$

$$d(P_{\text{S}})/dt = - \left[\mu_{\text{NF}}^{\text{max}} \frac{P_{\text{S}}}{P_{\text{S}} + P_{\text{H}}} \cdot \frac{\text{NF}}{R_{\text{ORG}}} \right] - \left[\mu_{\text{O}}^{\text{max}} \cdot \text{minimum} \left\{ \frac{P_{\text{S}}}{P_{\text{S}} + P_{\text{H}}}, \frac{N_{\text{S}}}{N_{\text{S}} + N_{\text{H}}} \right\} \cdot \frac{\text{O}}{R_{\text{ORG}}} \right] + \left[M \cdot \text{SR} \cdot \frac{\text{NF}}{R_{\text{ORG}}} \right] + \left[M \cdot \text{SR} \cdot \frac{\text{O}}{R_{\text{ORG}}} \right] + \left[\frac{K \cdot (P_{\text{D}} - P_{\text{S}})}{\text{SD}} \right] + \left[\frac{\text{RP}}{\text{SD}} \right] \quad (3)$$

$$d(N_{\text{S}})/dt = - \left[\mu_{\text{O}}^{\text{max}} \cdot \text{minimum} \left\{ \frac{P_{\text{S}}}{P_{\text{S}} + P_{\text{H}}}, \frac{N_{\text{S}}}{N_{\text{S}} + N_{\text{H}}} \right\} \cdot \text{O} \right] + [M \cdot (\text{SR} - 0.75 \text{ DN}) \cdot \text{NF}] + [M \cdot (\text{SR} - 0.75 \text{ DN}) \cdot \text{O}] + \left[\frac{K \cdot (N_{\text{D}} - N_{\text{S}})}{\text{SD}} \right] + \left[\frac{\text{RN} + \text{AN}}{\text{SD}} \right] \quad (4)$$

$$d(P_{\text{D}})/dt = + \left[M \cdot \text{DR} \cdot \frac{\text{NF}}{R_{\text{ORG}}} \cdot \frac{\text{SD}}{\text{DD}} \right] + \left[M \cdot \text{DR} \cdot \frac{\text{O}}{R_{\text{ORG}}} \cdot \frac{\text{SD}}{\text{DD}} \right] - \left[\frac{K \cdot (P_{\text{D}} - P_{\text{S}})}{\text{DD}} \right] \quad (5)$$

$$d(N_{\text{D}})/dt = + \left[M \cdot (\text{DR} - 0.25 \text{ DN}) \cdot \text{NF} \cdot \frac{\text{SD}}{\text{DD}} \right] + \left[M \cdot (\text{DR} - 0.25 \text{ DN}) \cdot \text{O} \cdot \frac{\text{SD}}{\text{DD}} \right] - \left[\frac{K \cdot (N_{\text{D}} - N_{\text{S}})}{\text{DD}} \right] \quad (6)$$

The P_{S} and P_{D} equations are written in units of mol P m^{-3} ; all other equations are in units of mol N m^{-3} .

Box 2 Model steady-state concentrations, nitrogen fluxes, and model solutions

Steady-state concentrations and nitrogen fluxes

Variable or nitrogen flux	Units	Model value	Literature values
NF, Nitrogen-fixers (standing stock)	Tg N	1	
(O + NF), Total phytoplankton (standing stock)	Tg N	110	115 (ref. 43), 150 (ref. 44), 105 (ref. 31)
P_S , Surface $[PO_4^{3-}]^*$	$\mu\text{mol kg}^{-1}$	0.15	0.50 (ref. 45)
N_S , Surface $[NO_3^-]^*$	$\mu\text{mol kg}^{-1}$	2.0	5.0 (ref. 45)
P_D , Deep $[PO_4^{3-}]$	$\mu\text{mol kg}^{-1}$	1.75	2.0 (ref. 45)
N_D , Surface $[NO_3^-]$	$\mu\text{mol kg}^{-1}$	26	30 (ref. 45)
R_S , Surface $[NO_3^-]:[PO_4^{3-}]$ ratio	(mol/mol)	13	~10 (ref. 9), ~10 (ref. 10)
R_D , Deep $[NO_3^-]:[PO_4^{3-}]$ ratio	(mol/mol)	14.6	~15 (ref. 9), ~14.6 (ref. 10)
Total (net) primary production (TPP)	Tg $N\text{ yr}^{-1}$	8,109	6,400–8,000 (ref. 44), 7,900–8,800 (ref. 46), 7,900 (ref. 18)
Surface regeneration	Tg $N\text{ yr}^{-1}$	7,612	75–97% TPP (Box 1)
Deep regeneration	Tg $N\text{ yr}^{-1}$	359	
Sinking flux (500 m)	Tg $N\text{ yr}^{-1}$	405	3–25% TPP (Box 1), 10% TPP (ref. 31)
Mixing flux (into surface layer)	Tg $N\text{ yr}^{-1}$	359	8% TPP (ref. 31)
Sedimentation flux	Tg $N\text{ yr}^{-1}$	16	0.04–1.0% TPP (Box 1)
N_2 -fixation flux	Tg $N\text{ yr}^{-1}$	69	80 (ref. 47), 10–130 (ref. 19), 19–66 (ref. 18), 30 (ref. 31), 75–175 (ref. 48)
Denitrification flux	Tg $N\text{ yr}^{-1}$	122	16–250 (Box 1)
River input	Tg $N\text{ yr}^{-1}$	30	$(2.6\text{--}8.4) \times 10^{-3} \text{ mol N m}^{-2} \text{ yr}^{-1} = 13\text{--}42$ (Box 1)
Atmospheric input (net)	Tg $N\text{ yr}^{-1}$	38	$(0.1\text{--}11.0) \times 10^{-3} \text{ mol N m}^{-2} \text{ yr}^{-1} = 0\text{--}56$ (Box 1)

*Nutrient concentrations vary over a large range in the upper 500 m, and with region (latitude). The values quoted here are very approximate averages.

Analytical solutions

$$P_S = \left[\frac{P_H}{(\mu'_{NF}/M) - 1} \right] \quad (7)$$

$$N_S = \left[\frac{N_H}{(\mu'_O/M) - 1} \right] \quad (8)$$

$$O = \left(\frac{1}{M \cdot SD} \right) \times \left[\left(\frac{RP \cdot R_{ORG} \cdot (1 - SF - DN)}{SF} \right) + (RN + AN) \right] \quad (9)$$

$$NF = \left(\frac{1}{M \cdot SD} \right) \times \left[\left(\frac{RP \cdot R_{ORG} \cdot (SF + DN)}{SF} \right) - (RN + AN) \right] \quad (10)$$

$$P_D = P_S + \left[\frac{RP \cdot (1 - SR - SF)}{K \cdot SF} \right] \quad (11)$$

$$= \left[\frac{P_H}{(\mu'_{NF}/M) - 1} \right] + \left[\frac{RP \cdot (1 - SR - SF)}{K \cdot SF} \right]$$

$$N_D = N_S + \left[\frac{RP \cdot R_{ORG} \cdot (1 - SR - SF - 0.25DN)}{K \cdot SF} \right] \quad (12)$$

$$= \left[\frac{N_H}{(\mu'_O/M) - 1} \right] + \left[\frac{RP \cdot R_{ORG} \cdot (1 - SR - SF - 0.25DN)}{K \cdot SF} \right]$$

$$R_S = \frac{N_S}{P_S} = \left(\frac{N_H}{P_H} \right) \cdot \left[\frac{\mu'_{NF} - M}{\mu'_O - M} \right] \quad (13)$$

$$R_D = \frac{N_D}{P_D} = \left\{ \frac{\left[\frac{R_{ORG} \cdot RP \cdot (1 - SR - SF - 0.25DN)}{K \cdot SF} \right] + \left[\frac{M \cdot N_H}{\mu'_O - M} \right]}{\left[\frac{RP \cdot (1 - SR - SF)}{K \cdot SF} \right] + \left[\frac{M \cdot P_H}{\mu'_{NF} - M} \right]} \right\} \quad (14)$$

$$TPP = \frac{RP \cdot R_{ORG}}{SF} \times 362 \times 10^{12} \times 14 \times 10^{-12} \quad (15)$$

R_S is the $[NO_3^-]:[PO_4^{3-}]$ ratio in the surface ocean, and R_D is the $[NO_3^-]:[PO_4^{3-}]$ ratio in the deep ocean. TPP is in units of Tg $N\text{ yr}^{-1}$.

deep and surface waters, and loss of reactive nitrogen due to denitrification. Denitrification is represented simply as a fixed proportion (1.5%) of the total remineralization, and is assumed to occur predominantly in the upper 500 m (75%), but also partly at greater depth (25%), for instance in sediments^{23,24}.

Model solutions

This model contains a mechanism providing an input of reactive nitrogen (by nitrogen fixation) when that reactive nitrogen is scarce relative to phosphate. The purpose of the model is to examine whether this mechanism, arising as a result of the ecological competition encoded in equations (1) and (2), can tie the nitrogen and phosphorus cycles together and thereby control the nitrogen cycle. One test of this is whether the model converges to a steady state when run, and if so whether that steady-state solution has reasonable concentrations and fluxes.

Equations (1)–(6) provide six differential equations in six separate variables. In the steady state, the differential with respect to time equals 0 in each case, and this provides six simultaneous equations in six variables, which can be solved analytically to give the steady-state solutions shown in Box 2. These analytical solutions are based on the understanding that nitrate must be more limiting than phosphate at the surface in the steady state, since otherwise NF and O (defined in Box 2) cannot simultaneously be balanced. Substituting the parameter values from Box 1 into equations (7)–

(14) gives the analytical steady-state solutions shown in Box 2.

This analytical steady-state solution was also found to be attainable in practice by numerical simulation using a fourth-order Runge–Kutta method, as shown for one particular set of initial conditions in Fig. 4a–d. Convergence occurs from a wide variety of initial states (Fig. 4e). Box 2 shows a good correspondence between both (1) steady-state concentrations and (2) steady-state global annual fluxes of nitrogen, and estimates, where available, of those concentrations and fluxes as they occur in nature. Parameter values such as N_H , P_H , μ'_O and μ'_{NF} could have been tuned to produce a better fit, but that would be to miss the point of this simple two-box model, which is intended only to explain the most important, first-order features of the N and P concentrations and ratios, which it is able to do. Despite organic matter with an N:P ratio of 16 and river inputs with a ratio of 30, the model converges to a steady state in which deep water has an $[NO_3^-]:[PO_4^{3-}]$ ratio of 14.6, and surface water has a ratio of 13.0.

Robust regulation of the nitrogen cycle

The model runs (Fig. 4a–e) show that the model converges to the analytically expected final state from all 10 randomly chosen sets of initial conditions. Other model runs (not shown) demonstrate that the system is also able to return to equilibrium after various types of perturbation: the model regulation of the nitrogen cycle is both homeostatic and robust.

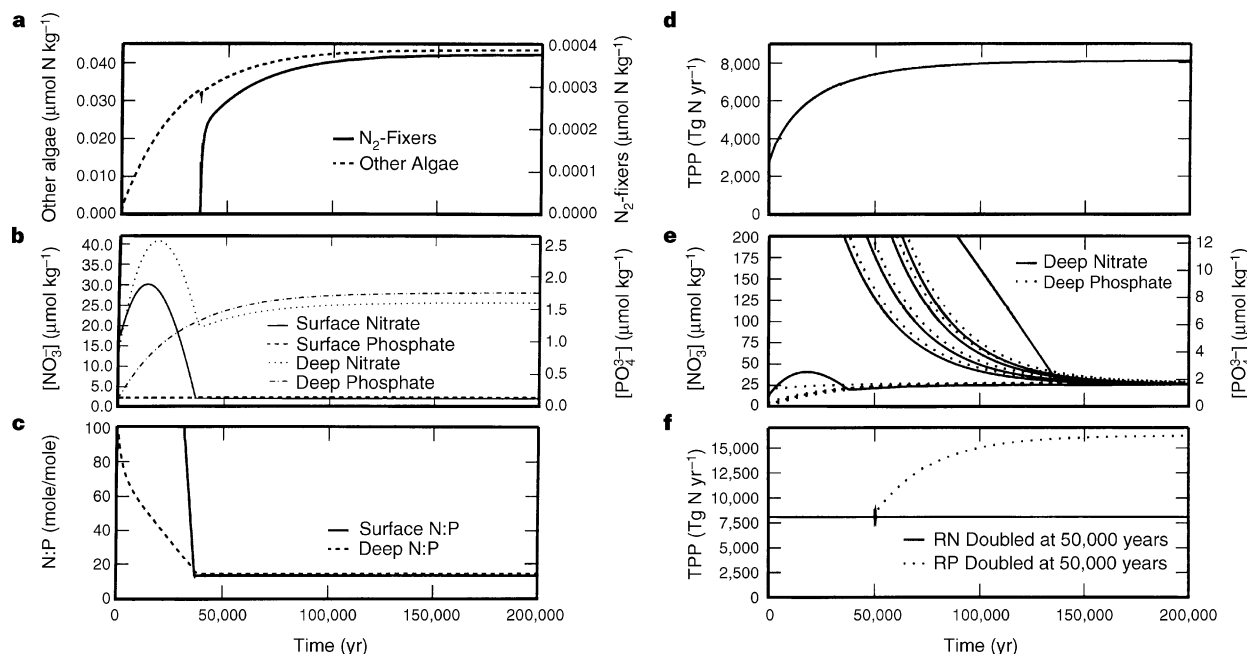


Figure 4 Results of model runs. **a–d**, Convergence to the steady state from an example initial state (randomly determined) in which $[NF] = 0.00013 \mu\text{mol N kg}^{-1}$, $[O] = 0.45 \mu\text{mol N kg}^{-1}$, $[P_S] = 0.0017 \mu\text{mol kg}^{-1}$, $[N_S] = 94 \mu\text{mol kg}^{-1}$, $[P_D] = 0.13 \mu\text{mol kg}^{-1}$, and $[N_D] = 0.48 \mu\text{mol kg}^{-1}$, giving an initial ocean average of $0.11 \mu\text{mol P kg}^{-1}$, $13 \mu\text{mol N kg}^{-1}$, $[NO_3^-]:[PO_4^{3-}]$ of 120. The final (equilibrium) state agreed with that expected from equations (7)–(14). Phytoplankton productivity was depressed by the initial low P availability, leading to small burial losses and increasing P until a balance of inputs (river) and outputs (burial) was achieved. At the same time, the large initial excess of nitrate depressed N₂-fixation, eventually reducing $[NO_3^-]:[PO_4^{3-}]$ until the $[NO_3^-]:[PO_4^{3-}]$ set-point was also reached. The model dynamics show two homeostats acting in parallel, the first regulating total nutrient levels via automatic adjustments in productivity, the second regulating $[NO_3^-]:[PO_4^{3-}]$ via automatic adjustments in the relative abundance of nitrogen-fixers. **e**, Convergence of simulation (only $[P_D]$ and $[N_D]$ are shown) to a steady

state from ten random initial states, each chosen by setting the six variables to their steady-state values multiplied or divided by a random number in the range 1–100. Some runs required minimum thresholds to protect $[NF]$ or $[O]$ under prolonged unfavourable conditions, or required density-dependent mortality to damp out oscillations. **f**, Response of TPP to a doubling of the river input of nitrate (solid line, $RN \rightarrow 12.0 \text{ mol P m}^{-2} \text{ yr}^{-1}$) or phosphate (dotted line, $RP \rightarrow 0.4 \text{ mol P m}^{-2} \text{ yr}^{-1}$). The changes in riverine nutrient inputs were first imposed at 50,000 years and then maintained to the end of the runs. Close examination of the model output at ~50,000 years shows that for both model runs nitrate is always the most limiting nutrient in surface waters (the PLN), with no interruptions, even though the size of the river input of phosphate is always controlling productivity (the ULN). Note the different axis scales in **a** for $[NF]$ and $[O]$, and in **b** and **e** for $[NO_3^-]$ and $[PO_4^{3-}]$.

Numerous sensitivity analyses (see Supplementary Information) showed clearly that the regulation of the N cycle in the model is not dependent on a finely tuned set of parameter values, and is also not affected by most assumptions about the processes involved. The regulation arises in a robust fashion from the thermostat-like N₂-fixation flux, which changes so as to counteract any deviations in the $[NO_3^-]:[PO_4^{3-}]$ ratio from the equilibrium 'set-point' value for the ocean.

Nitrate versus phosphate limitation

A distinction between ultimate and proximate limiting nutrients was proposed earlier. By incorporating an ecological competition between the two relevant types of phytoplankton, a geochemical model has been produced which predicts that the PLN in the oceans is nitrate (as observed by biologists), while the ULN is phosphate (as predicted by geochemists). This model can explain not only why the $[NO_3^-]$ versus $[PO_4^{3-}]$ scatter-plot average trend intersects the axes near the origin, but also why it intersects a short distance away from the origin along the P-axis (Fig. 1), that is, why there is proximate nitrate limitation of surface waters. This is among several explanations for proximate nitrogen limitation previously considered⁴⁹.

An increase in the river delivery of nitrate has no long-term effect on productivity (Fig. 4f). An increase in the river delivery of phosphate, on the other hand, causes a sustained and proportionate increase in productivity. This is in line with the analytical solution for TPP at steady state ($= M(NF + O)$), which is proportional to the riverine input of phosphorus (RP) but independent of inputs of nitrogen (RN and AN) (equation (15)).

How can the PLN and ULN be different in this way? At the steady state there is a nitrate deficit ($[NO_3^-] < 16[PO_4^{3-}]$) at the surface and a small population of nitrogen-fixers (<1% of total phytoplankton) constantly adding new nitrate. Injection of extra nitrate in the river supply increases production in the very short term, but also simultaneously diminishes the niche for nitrogen-fixers (by shifting the environment closer to the top of Fig. 3c), causing fewer N₂-fixers and hence reduced new-nitrogen input. Eventually a new steady state is reached in which the increased river input of nitrate is offset by a smaller N₂-fixation input from a smaller population of N₂-fixers. The total nitrate input and TPP will eventually be as before.

The injection of extra phosphate, on the other hand, increases the nitrate deficit at the surface, causing a steady state with a larger population of nitrogen-fixers, a greater total N-input than before (matching the greater P-input), and greater TPP.

A surface ocean where N and P are equally limiting (a 'Redfield ocean') is unlikely to be a steady state, because N₂-fixers would be out-competed due to their lower maximum growth rate (representing nitrogenase-building costs and energetic costs of breaking N₂ triple bonds). They would consequently decrease in number and slowly disappear from the system, until the resultant reduction in new nitrate input once more produced a nitrate-limited surface ocean. In the steady state reactive N has to be more limiting to compensate for the N₂-fixers' inherent disadvantage otherwise; that is, N-limitation must be more severe than P-limitation, and so N must be the PLN.

Equation (13) shows that the exact numerical value of the steady-

state solution for surface $[\text{NO}_3^-]:[\text{PO}_4^{3-}]$ depends strongly on the values of poorly-known biological parameters such as P_H and N_H . However, nitrate limitation is robustly predicted to be more severe than phosphate limitation in the steady state, because equations (1) and (2) cannot otherwise simultaneously be balanced.

Is the ocean losing nitrate?

It has been argued¹³ that the ocean may slowly be losing nitrate because denitrification is currently removing reactive nitrogen more rapidly than N_2 -fixation is importing it. However, a review of all fluxes in the nitrogen cycle (Box 2), including river and atmospheric inputs and burial losses, shows that an imbalance between denitrification and nitrogen fixation fluxes is in fact compatible with a balance between total inputs and total outputs of reactive nitrogen, at least to the accuracy with which we can currently estimate the different fluxes. In addition, the model demonstrates that an overall ocean nitrate deficit is compatible with a system in steady state, and does not necessarily imply that the ocean is losing nitrate.

The hypothesis developed here and quantitatively examined in this model contrasts with a recent proposal¹⁴ that the widespread ocean nitrate deficit is indicative of continuing nitrate loss from the oceans, caused by iron limitation of N_2 -fixers in the open ocean. An increased supply of iron to the oceans (for example, during drier glacial times) would then favour nitrogen fixation and would cause the ocean nitrate deficit to be removed. This proposal is in line with the high iron requirement for building nitrogenase⁸, but it also implicitly assumes that the steady state of the ocean (when iron is abundant enough so that N_2 -fixers are not limited by it) is $[\text{NO}_3^-]:[\text{PO}_4^{3-}] = 16$.

The results I report here contest that assumption. Nitrate deficits are accounted for parsimoniously, without recourse to trace-metal effects, with my model simultaneously giving reasonable outputs in terms of steady-state fluxes and standing stocks. My model predicts that changes in N_2 -fixation may well have compensated for any glacial–interglacial changes in denitrification, and that the ocean does not need to have lost or gained nitrate (relative to phosphate) over glacial–interglacial cycles. The situation of $\text{PLN} = \text{nitrate}$, $\text{ULN} = \text{phosphate}$ may have held true continuously for geological timescales.

However, although my model predicts a baseline nitrate deficit that will persist even when there is a surplus of iron in the oceans, it does not rule out shortages of iron causing nitrate deficits greater than the baseline.

Implications and testability

These findings have possible implications for management of nutrient fluxes to coastal waters and semi-enclosed seas, especially where mixing between surface and deep waters is more rapid than for the global ocean as a whole. If the model representation of nutrient limitation of phytoplankton growth is correct, then it is phosphates, not nitrates, which should be removed from waste water in order to reduce eutrophication. Removal of nitrates in the river supply should lead to increased nitrogen fixation, no significant effect on final nitrate concentrations, and no significant effect on eutrophication. This work therefore supports the conclusion²⁵ that “the recommendation to remove nitrate, from waste water and influx to rivers, in order to counteract eutrophication, has no tenable scientific basis”, despite “enrichment experiments indicating nitrogen limitation”. Nitrate removal may however be preferable in order to limit ‘end-pipe’, rather than larger-scale, eutrophication effects.

My model allows us to understand findings such as that from a 23-year study of Narragansett Bay, which showed a positive correlation between plankton biomass and phosphorus inputs, but no such correlation with nitrogen inputs²⁶, and similar results (also some evidence of P-stimulation of fisheries) from long-term studies of the

North Sea²⁷ and of the Dutch Wadden Sea²⁸.

Ways of testing this model and its conclusions include: (1) improved experimental determination of average values of P_H , N_H and the relative difference between μ'_O and μ'_{NF} ; and (2) comparison of N and P cycling and nitrate deficits in iron-rich versus iron-poor regions in the present-day ocean, to see whether nitrate deficits persist when iron concentrations are high.

Resolving the debate

My model produces robust and homeostatic control of nitrate and phosphate concentrations, opposing deviations from a steady state that agrees well with estimated real-world fluxes and standing stocks. The ecological competition between nitrogen-fixing and non-nitrogen-fixing algae controls nitrate levels, pushing nitrate to concentrations continually slightly less than 16 times the phosphate concentration. This gives rise to a nitrogen homeostat (‘nitrostat’), with the ocean acting like a giant thermostat, but automatically regulating its own $[\text{NO}_3^-]:[\text{PO}_4^{3-}]$ ratio not its temperature. When the value of this ratio in the oceans falls too low, nitrogen fixation inputs increase, restoring the ratio. When the $[\text{NO}_3^-]:[\text{PO}_4^{3-}]$ ratio rises too high, nitrogen fixation inputs decrease, again restoring the ratio. The importance of nitrogen-fixers is strongly supported by this modelling study, which shows how their influence is exerted.

The relatively small size of the current marine nitrogen-fixation flux is shown to be compatible with its role as the long-term controller of ocean nitrate content. Nitrate is the proximate limiting nutrient in surface waters; that is, the most limiting to instantaneous growth according to Liebig’s law. Phosphorus, however, is predicted to be the ultimate limiting nutrient, whose rate of supply simultaneously regulates total ocean productivity. My model can thereby resolve the long-standing debate over the relative importance of nitrate and phosphate for ocean productivity. □

Received 30 November 1998; accepted 1 June 1999.

1. Redfield, A. C. in *James Johnston Memorial Volume* 176–192 (Liverpool Univ. Press, 1934).
2. Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Lamont-Doherty Geological Observatory, New York, 1982).
3. Libes, S. M. *An Introduction to Marine Biogeochemistry* (Wiley, New York, 1992).
4. Schindler, D. W. Evolution of phosphorus limitation in lakes. *Science* **195**, 260–262 (1977).
5. Schindler, D. W. Whole-lake eutrophication experiments with phosphorus, nitrogen and carbon. *Int. Ver. Theor. Angew. Limnol. Verh.* **19**, 3221–3231 (1975).
6. Smith, V. H. Low nitrogen to phosphorus ratios favor dominance by blue-green algae in lake phytoplankton. *Science* **221**, 669–671 (1983).
7. Hecky, R. E. & Kilham, P. Nutrient limitation of phytoplankton in freshwater and marine environments: a review of recent evidence on the effects of enrichment. *Limnol. Oceanogr.* **33**, 796–822 (1988).
8. Howarth, R. W., Marino, R. & Cole, J. J. Nitrogen fixation in freshwater, estuarine and marine ecosystems. 2. Biogeochemical controls. *Limnol. Oceanogr.* **33**, 688–701 (1988).
9. Geochemical Oceans Section Study (GEOSECS) data set. (cited 1996) (<http://ingrid.ligo.columbia.edu/SOURCES/GEOSECS>).
10. Tyrrell, T. & Law, C. S. Low nitrate:phosphate ratios in the global ocean. *Nature* **387**, 793–796 (1997); Low nitrate: phosphate ocean ratios corrected. *Nature* **393**, 318 (1998).
11. Ryther, J. G. & Dunstan, W. M. Nitrogen, phosphorus and eutrophication in the coastal marine environment. *Science* **171**, 1008–1013 (1971).
12. Codispoti, L. A. in *Productivity of the Ocean: Present and Past* (eds Berger, W. H., Smetacek, V. S. & Wefer, G.) 377–394 (Dahlem Conf., Wiley, Chichester, UK, 1989).
13. Codispoti, L. A. Is the ocean losing nitrate? *Nature* **376**, 724 (1995).
14. Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological CO_2 pump in the ocean. *Nature* **387**, 272–275 (1997).
15. Smith, S. V. Phosphorus versus nitrogen limitation in the marine environment. *Limnol. Oceanogr.* **29**, 1149–1160 (1984).
16. Broecker, W. S. A kinetic model for the composition of sea water. *Quat. Res.* **1**, 188–207 (1971).
17. Jahnke, R. A. in *Global Biogeochemical Cycles* (eds Butcher, S. S., Charlson, R. J., Orians, G. O. & Wolfe, G. V.) 301–315 (Academic, London, 1992).
18. Mackenzie, F. T., Ver, L. M., Sabine, C., Lane, M. & Lerman, A. in *Interactions of C, N, P and S Biogeochemical Cycles and Global Change* (eds Wollast, R., Mackenzie, F. T. & Chou, L.) 1–61 (NATO ASI Ser. Vol. 14, Springer, Berlin, 1993).
19. Jaffe, D. A. in *Global Biogeochemical Cycles* (eds Butcher, S. S., Charlson, R. J., Orians, G. O. & Wolfe, G. V.) 263–284 (Academic, London, 1992).
20. Fasham, M. J. R., Ducklow, H. W. & McKelvie, S. M. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *J. Mar. Res.* **48**, 591–639 (1990).
21. Kimmerer, W. J., Smith, S. V. & Hollibaugh, J. T. A simple heuristic model of nutrient cycling in an estuary. *Estuar. Coast. Shelf Sci.* **37**, 145–159 (1993).
22. Banse, K. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 409–439 (Plenum, New York, 1992).
23. Christensen, J. P., Townsend, D. W. & Montoya, J. P. Water column nutrients and sedimentary denitrification in the Gulf of Maine. *Cont. Shelf Res.* **16**, 489–515 (1996).
24. Anderson, L. A. & Sarmiento, J. L. Redfield ratios of remineralisation determined by nutrient data analysis. *Glob. Biogeochem. Cycles* **8**, 65–80 (1994).
25. Söderström, J. The significance of observed nutrient concentrations in the discussion about nitrogen and phosphorus as limiting nutrients for the primary carbon flux in coastal water ecosystems. *Sarsia*

- 81, 81–96 (1996).
26. Smayda, T. J. Narragansett Bay: variability and change in environment and phytoplankton dynamics over a 23-year period. (Abstr.) *Estuaries* **8**, 86 (1985).
27. Boddeke, R. & Hagel, P. in *Proc. World Fisheries Congress, Theme 1* (ed. Armantrout, N. B.) 290–315 (Oxford & IBH, PVT, New Delhi, 1994).
28. Nanninga, H. J., Gieskes, W. W. C. & Wolff, W. J. *Fish Production in the North Sea and the Influence of Nitrogen and Phosphorus* (Literature Study, Dept of Marine Biology, Univ. Groningen, 1997).
29. Suess, E. Particulate organic carbon flux in the oceans—surface productivity and oxygen utilisation. *Nature* **288**, 260–263 (1980).
30. Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. W. VERTEX: carbon cycling in the northeast Pacific. *Deep-Sea Res.* **34**, 267–285 (1987).
31. Schlesinger, W. H. *Biogeochemistry: an Analysis of Global Change* (Academic, San Diego, 1991).
32. Meybeck, M. in *Interactions of C, N, P and S Biogeochemical Cycles and Global Change* (eds Wollast, R., Mackenzie, F. T. & Chou, L.) 163–193 (NATO ASI Ser. Vol. 14, Springer, Berlin, 1993).
33. Berner, E. K. & Berner, R. A. *Global Environment: Water, Air, and Geochemical Cycles* (Prentice-Hall, NJ, 1996).
34. Cornell, S., Rendell, A. & Jickells, T. Atmospheric inputs of dissolved organic nitrogen to the oceans. *Nature* **376**, 243–246 (1995).
35. Duce, R. A. *et al.* The atmospheric input of trace species to the world ocean. *Glob. Biogeochem. Cycles* **5**, 193–259 (1991).
36. Copin-Montegut, C. & Copin-Montegut, G. Stoichiometry of carbon, nitrogen, and phosphorus in marine particulate matter. *Deep-Sea Res.* **30**, 31–46 (1983).
37. Rios, A. F., Fraga, F. & Pérez, F. F. Estimation of coefficients for the calculation of “NO”, “PO” and “CO”, starting from the elemental composition of natural phytoplankton. *Scientia Marina* **53**, 779–784 (1989).
38. Furnas, M. J. *In situ* growth rates of marine phytoplankton: approaches to measurement, community and species growth rates. *J. Plankt. Res.* **12**, 1117–1151 (1990).
39. McAllister, C. D., Shah, N. & Strickland, J. D. H. Marine phytoplankton photosynthesis as a function of light intensity: a comparison of methods. *J. Fish. Res. Board Can.* **21**, 159–181 (1964).
40. Davies, A. G. & Sleep, J. A. The photosynthetic response of nutrient-depleted dilute cultures of *Skeletonema costatum* to pulses of ammonium and nitrate; the importance of phosphate. *J. Plankt. Res.* **11**, 141–164 (1981).
41. Eppley, R. W., Rogers, J. N. & McCarthy, J. J. Half-saturation constants for uptake of nitrate and ammonium by marine phytoplankton. *Limnol. Oceanogr.* **14**, 912–920 (1969).
42. Goldman, J. C. & Glibert, P. M. in *Nitrogen in the Marine Environment* (eds Capone, D. G. & Carpenter, E. J.) 233–274 (Academic, New York, 1983).
43. Barnes, R. S. K. & Hughes, R. N. *An Introduction to Marine Ecology* 2nd edn (Blackwell Science, Oxford, 1988).
44. Antoine, D., André, J.-M. & Morel, A. Oceanic primary production, 2. Estimation at global scale from satellite (Coastal Zone Colour Scanner) chlorophyll. *Glob. Biogeochem. Cycles* **10**, 57–69 (1996).
45. Levitus, S., Conkright, M. E., Reid, J. L., Najjar, R. G. & Mantyla, A. Distribution of nitrate, phosphate and silicate in the world oceans. *Prog. Oceanogr.* **31**, 245–273 (1993).
46. Longhurst, A., Sathyendranath, S., Platt, T. & Caverhill, C. An estimate of global primary production in the ocean from satellite radiometer data. *J. Plank. Res.* **17**, 1245–1271 (1995).
47. Capone, D. G., Zehr, J. P., Paele, H. W., Bergman, B. & Carpenter, E. J. *Trichodesmium*, a globally significant cyanobacterium. *Science* **276**, 1221–1229 (1997).
48. Gruber, N. & Sarmiento, J. L. Global patterns of marine nitrogen fixation and denitrification. *Glob. Biogeochem. Cycles* **11**, 235–266 (1997).
49. Vitousek, P. M. & Howarth, R. W. Nitrogen limitation on land and in the sea: How can it occur? *Biogeochemistry* **13**, 87–115 (1991).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements. I thank E. Marañón, T. Lenton, A. Martin, D. Wolf-Gladrow, T. Anderson, H. Nanninga, J. Shepherd, P. Holligan, P. Herman, W. Stolte, C. Zonneveld, A. Taylor, A. Watson, L. Partridge, W. Barkmann, S. Smith and R. Toggweiler for discussions and comments on the manuscript, and A. Brice for assistance with computing.

Correspondence and requests for materials should be addressed to the author (e-mail: T.Tyrrell@soc.soton.ac.uk).

The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*

S. Bowman, D. Lawson, D. Basham, D. Brown, T. Chillingworth, C. M. Churcher, A. Craig*, R. M. Davies, K. Devlin, T. Feltwell, S. Gentles, R. Gwilliam, N. Hamlin, D. Harris, S. Holroyd, T. Hornsby, P. Horrocks*, K. Jagels, B. Jassal, S. Kyes*, J. McLean, S. Moule, K. Mungall, L. Murphy, K. Oliver, M. A. Quail, M.-A. Rajandream, S. Rutter, J. Skelton, R. Squares, S. Squares, J. E. Sulston, S. Whitehead, J. R. Woodward, C. Newbold* & B. G. Barrell

Pathogen Sequencing Unit, Sanger Centre, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK

* Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, UK

Analysis of *Plasmodium falciparum* chromosome 3, and comparison with chromosome 2, highlights novel features of chromosome organization and gene structure. The sub-telomeric regions of chromosome 3 show a conserved order of features, including repetitive DNA sequences, members of multigene families involved in pathogenesis and antigenic variation, a number of conserved pseudogenes, and several genes of unknown function. A putative centromere has been identified that has a core region of about 2 kilobases with an extremely high (adenine + thymidine) composition and arrays of tandem repeats. We have predicted 215 protein-coding genes and two transfer RNA genes in the 1,060,106-base-pair chromosome sequence. The predicted protein-coding genes can be divided into three main classes: 52.6% are not spliced, 45.1% have a large exon with short additional 5' or 3' exons, and 2.3% have a multiple exon structure more typical of higher eukaryotes.

Malaria remains a major burden to human health in tropical and subtropical areas. In Africa alone, more than one million children under five die from the disease each year¹. Although four members of the *Plasmodium* genus normally infect humans, nearly all deaths are attributable to a single parasite species, *P. falciparum*. The severity of disease caused by this species results primarily from its ability to modify the surface of infected red blood cells by inserting parasite proteins. Parasitized erythrocytes can bind to host endothelial cells, a process called cytoadherence, leading in some cases to their accumulation in specific organs such as the brain and to the development of cerebral malaria². Current approaches to malaria control and treatment rely on measures such as insecticide-impregnated bed nets and chemotherapy. Although considerable resources have been devoted to the development of a vaccine, no effective immunization regime so far exists. Moreover, the rapid spread of resistance to existing and new antimalarial drugs means that in some areas of the world, particularly southeast Asia, reliable prophylaxis is not possible, making treatment difficult³.

In response to these problems, the Malaria Genome Sequencing Consortium⁴ was established to sequence the entire genome of *P. falciparum* as a collaborative venture. In less than a year, the consortium aims to generate almost all of the *P. falciparum* genome sequence in unfinished form, making nearly its entire gene complement accessible for malaria researchers. Focus is already shifting to the development and implementation of whole genome approaches to drug development and vaccine target identification.

Sequencing strategy

P. falciparum has a nuclear genome of around 30 megabases (Mb) divided between 14 chromosomes, which range in size from 0.7 to 3.5 Mb. Sequencing this genome presents significant technical difficulties, mainly because of the biased nucleotide composition of its DNA, with an overall (A + T) content estimated at 82% (ref. 5). In general, fragments of DNA over 5 kilobases (kb) are unstable in *Escherichia coli*, so the large-insert bacterial clones commonly used as templates for sequencing are not available. However, most *P. falciparum* chromosomes can be resolved by pulsed-field gel

electrophoresis (PFGE), and some mapped yeast artificial chromosomes (YACs) exist. Chromosome 3 was sequenced using a whole chromosome 'shotgun' (WCS), an approach that was pioneered during the sequencing of *Saccharomyces cerevisiae* chromosome IX (ref. 6), and which was also used for *P. falciparum* chromosome 2 (ref. 7). Management of the WCS was facilitated by generating several-hundred sequence reads from each of the YACs^{8,9} comprising the minimal tiling path for this chromosome, and using these to place the WCS into a series of discrete bins. A strategy of sequencing YAC clones exclusively was discarded because of the high frequency of chimaerism, deletions and rearrangements inherent to YAC libraries¹⁰. In addition, although the YACs were purified through two pulsed-field gels¹¹, the level of contaminating *S. cerevisiae* sequence remained high owing to preferential cloning of yeast DNA.

Assembly validation

Because of the complexity associated with the assembly of an entire chromosome containing extensive repetitive regions, we confirmed

Figure 1 *P. falciparum* chromosome 3. Order and orientation of genes and predicted genes are shown. Exons are shown as coloured boxes with introns as linking lines. The two tRNA genes are shown as purple boxes with gene names shown in purple. Genes encoding proteins that have been characterized previously in *P. falciparum* have names shown in red. Predicted genes encoding proteins that have similarities to proteins currently unique to apicomplexan species are shown in fuchsia. Predicted genes encoding proteins with similarity to proteins in other eukaryotes for which some functional information is available are shown in yellow. Predicted genes encoding proteins with similarities to proteins of unknown function in other organisms are shown in orange. Predicted genes encoding proteins with similarity solely to bacterial proteins are shown in light blue. Genes encoding predicted proteins having similarity only within a defined protein domain are shown in grey. Predicted genes having no significant similarities are shown in dark green with those that have been confirmed by RT-PCR shown in light green. Pseudogenes are shown in dark blue. N-terminal signal sequences are represented by wavy lines. Repetitive telomeric sequences are shown as hatched boxes, and the location of the predicted centromere (put.CEN) is shown as a red box.

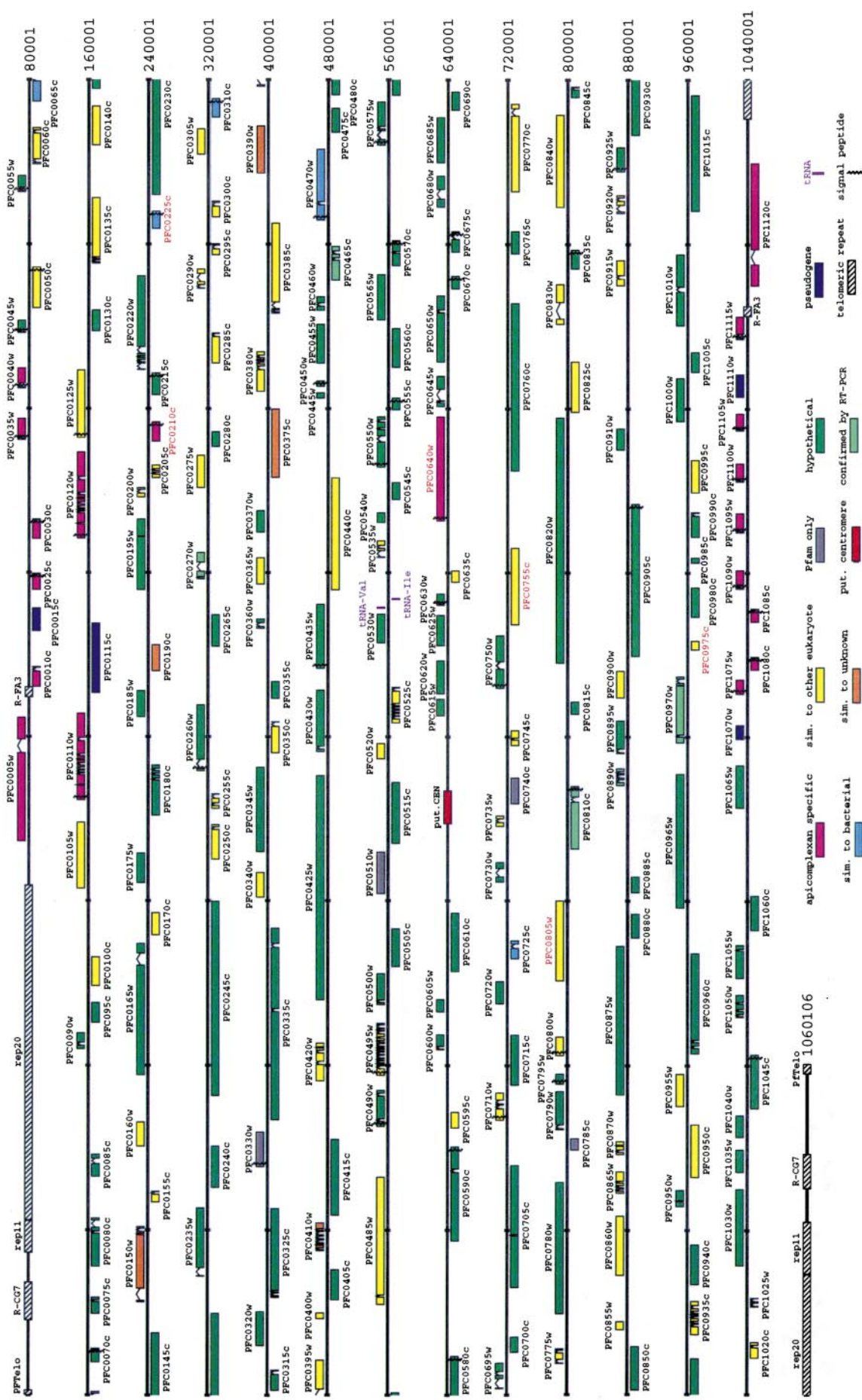


Table 1 Preliminary classification of chromosome-3 proteins

(a)			
Gene name	Description	Gene name	Description
Metabolism		Transport	
PFC0050C	Long chain fatty acid CoA ligase	PFC0125W	ABC transporter
PFC0170C	Lipoamide acyltransferase	PFC0725C	Formate transporter
PFC0275W	Glycerol-3-phosphate dehydrogenase	PFC0840W	E1-E2 P-type ATPase
PFC0395W	Asparagine synthetase	Cell surface	
PFC0710W	Inorganic pyrophosphatase	PFC0005W	<i>var</i> (3D7-varT3-1)
PFC0830W	Triosephosphate isomerase	PFC0010C	<i>rifin</i> (3D7-rifT3-1)
PFC0935C	<i>N</i> -glucosamine-1-phosphate transferase	PFC0025C	<i>stevor</i> (3D7-stevorT3-1)
PFC0950C	Acylaminoacyl peptidase (APH)	PFC0030C	<i>rifin</i> (3D7-rifT3-2)
PFC0995C	Acyl-CoA, cholesterol acyltransferase (ACAT)	PFC0035W	<i>rifin</i> (3D7-rifT3-3)
Cell growth, division, DNA synthesis		PFC0040W	<i>rifin</i> (3D7-rifT3-4)
PFC0305W	Putative microtubule binding protein	PFC0110W	Cytoadherence-linked asexual protein (3D7 clag-3.2)
PFC0340W	DNA polymerase delta, small subunit	PFC0120W	Cytoadherence-linked asexual protein (3D7 clag-3.1)
PFC0385C	Serine/threonine protein kinase	PFC0210C	Circumsporozoite (CS) protein
PFC0525C	Glycogen synthase kinase	PFC0640W	CTRP
PFC0595C	PP2A phosphatase	PFC0800W	Putative band 7 protein (stomatin)
PFC0755C	CDC2-related protein kinase	PFC1095W	<i>rifin</i> (3D7-rifT3-5)
PFC0770C	Kinesin-related protein	PFC1100W	<i>rifin</i> (3D7-rifT3-6)
PFC0860W	Kinesin-related protein	PFC1105C	<i>stevor</i> (3D7-stevorT3-2)
Transcription and post-transcriptional modification		PFC1115C	<i>rifin</i> (3D7-rifT3-7)
PFC00155C	DNA-directed RNA polymerase, 14K subunit	PFC1120C	<i>var</i> (3D7-varT3-2)
PFC0805W	DNA-directed RNA polymerase II, largest subunit	Intracellular trafficking	
PFC0825C	Cleavage and polyadenylation specificity factor	PFC0135C	Chromosome region maintenance protein
Protein synthesis		Cellular organization / biogenesis	
PFC0531C	tRNA-Val	PFC0920W	Histone H2A variant
PFC0532W	tRNA-Ile	Cell rescue, defence, death and ageing	
PFC0200W	60S ribosomal protein L44	PFC0205C	Glutaredoxin
PFC0225C	Elongation factor TS (EF-TS)	PFC0250C	AP endonuclease
PFC0290W	40S ribosomal protein S23	PFC0975C	Cyclophilin
PFC0295C	40S ribosomal protein S12	Unknown function	
PFC0300C	60S ribosomal protein L7	PFC0150W	Homologue of human protein KIAA0249
PFC0400W	60S acidic ribosomal protein P2	PFC0190C	Homologue of <i>D. melanogaster</i> PAST protein
PFC0470W	valyl-tRNA synthetase	PFC0375C	Homologue of <i>C. elegans</i> T08A11.2 protein
PFC0535W	60S ribosomal protein L26	PFC0390W	Putative homologue of <i>C. elegans</i> Y48E1C.2 protein
PFC0635C	Eukaryotic translation initiation factor 4E (EIF-4E)	PFC0410W	Putative homologue of rat brain YT51 protein
PFC0735W	40S ribosomal protein S15A	PFC1025C	Homologue of <i>C. elegans</i> F49C12.11 protein
PFC0775W	40S ribosomal protein S11	PFC1075W	Homologue of PFB0980W
PFC0870W	Elongation factor 1-beta	PFC1080C	Homologue of PFB0985C
PFC1020C	40S ribosomal protein S3A	PFC1085C	Homologue of PFB0990C
Protein destination		PFC1090W	Homologue of PFB0995W
PFC0140C	<i>N</i> -ethylmaleide-sensitive fusion protein NSF		
PFC0255C	Ubiquitin-conjugating enzyme E2		
PFC0285C	T-complex protein, beta subunit		
PFC0310C	ATP-dependent CLP protease		
PFC0350C	T-complex protein, Eta subunit		
PFC0495W	Aspartyl protease		
PFC0520W	26S proteasome regulatory subunit S14		
PFC0745C	Proteasome component C8 (macropain subunit 8)		
PFC0855W	Ubiquitin conjugating enzyme E2-17K		
PFC0900W	T-complex protein I, epsilon subunit		

(b)	
Pfam family	Gene name
Zinc-finger (C3HC4) protein	PFC0510W, PFC0740C
ATP-dependent RNA helicase	PFC0440C, PFC0915W, PFC0955C
PDZ domain protein	PFC0330W, PFC0785C
Serine/threonine protein kinase	PFC0060C, PFC0105W, PFC0485W
Guanine-nucleotide-binding protein	PFC100C, PFC0365W
Ankyrin repeat protein	PFC0160W
Calcium-dependent protein kinase	PFC0420W
Dual-specificity protein phosphatase	PFC0380W
Alpha beta hydrolase	PFC0065C
RNA-binding protein	PFC0865W

a. Gene names and predicted function of the product encoded are listed, with protein classification adapted from that used for *S. cerevisiae*. **b.** Proteins encoded on chromosome 3 that contain a domain identified in existing protein families, but which currently cannot be given a functional classification.

the chromosome 3 sequence by several independent methods. As double-stranded clones were used, with reads produced using both forward and reverse primers, the consistency of the read pairs generated was used as an initial assembly check. Reads derived from mapped YAC clones allowed confirmation of the colinearity of the chromosome assembly and the chromosome 3 YAC map⁸. In addition, the restriction enzyme map generated for this chromosome during the mapping project was used as further confirmation that the sequence had assembled correctly⁸. The order of mapped sequence-tagged site markers and simple sequence-length poly-

morphism microsatellite markers generated from the HB3xDd2 linkage segregation genetic map¹² were also confirmed in the final assembly. Finally, the DNA sequence correlated well with the restriction enzyme pattern for chromosome 3 generated by optical mapping¹³, with an average error of ~5%.

Analysis

The 1,060,106-bp chromosome 3 sequence encodes 215 predicted proteins and two tRNAs (Fig. 1), yielding a mean gene density of one predicted gene every 4.8 kb, which is similar to that in

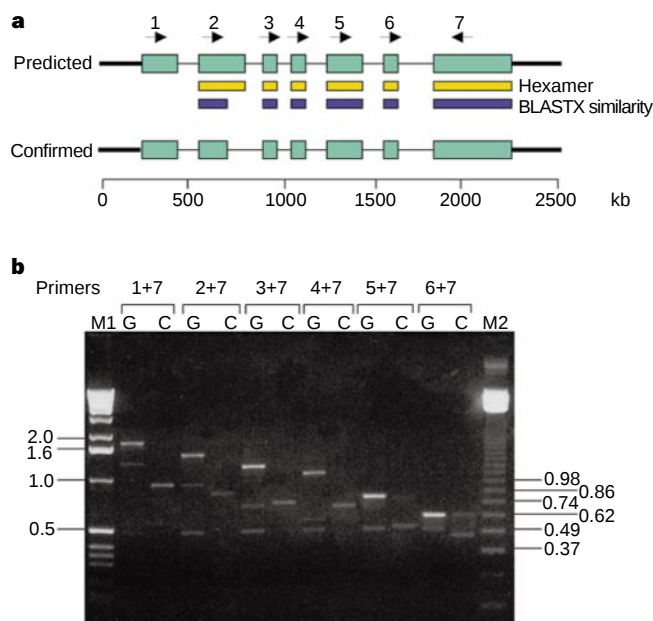


Figure 2 Splicing in the gene PFC0935C. **a**, The original gene prediction is shown, generated using hexamer and GENEFINDER, with exons drawn as coloured boxes and introns as thin lines, together with the supporting hexamer prediction of protein-coding potential and BLASTX similarity data. This is compared with the *in vivo* splicing of this gene, confirmed by RT-PCR and DNA sequencing. Exon 1 is not supported by either protein similarity or hexamer prediction, and exon 2 is incorrectly predicted by the hexamer program. Primers used for exon confirmation are shown as numbered arrows. **b**, A gel showing comparison of PCR products generated from genomic (G) and cDNA (C) template, with primer combinations shown above. Combinations 1 + 7 and 2 + 7 gave slightly smaller products than predicted, an observation confirmed by sequence data. DNA size markers, the 1-kb ladder (lane M1) and the 123-bp ladder (lane M2) are shown in kb.

Caenorhabditis elegans (one gene per 5 kb)¹⁴. The overall (A + T) composition of this chromosome is 80%, with the base composition of exons being 76.8% (A + T) and introns being 84.6% (A + T). Six of the genes on chromosome 3 had been characterized prior to this work. They encode circumsporozoite protein (CSP)^{15,16}, CS protein-TRAP-related protein (CTRP)¹⁷, RNA polymerase II largest subunit¹⁸, elongation factor-TS, CDC2-related protein kinase (EMBL accession no. EM:M86715; TREMBL accession no. TR:Q25028) and cyclophilin¹⁹, of which only three had been mapped to this chromosome^{15–18}. Ninety-four of the predicted proteins (43.7%) have significant similarity to existing database entries. Eighty-five (39.5%) have matches to eukaryotic proteins, many providing potential functional information. Nineteen of these (8.8%) are currently unique to *P. falciparum* or other apicomplexan parasites, but few proteins in this group have been functionally characterized. Five predicted proteins (2.3%) have significant similarity solely to bacterial proteins and are likely to be localized to the organelles of the parasite. In total, 27.4% of the predicted proteins on chromosome 3 are members of Pfam protein families²⁰. A further four predicted proteins contain discrete protein domains, but similarity does not extend further than the domain identified (Fig. 1). A preliminary functional classification of the proteins encoded on chromosome 3 is shown in Table 1.

Most of the predicted proteins (202 proteins; 94.0%) contain low-complexity, non-globular regions, as defined by the seg program²¹. However, the percentage of residues defined as low-complexity is 21.6%, indicative of the small size of many of these regions. Such regions have been previously reported⁷ and are often polymorphic between parasite isolates in both housekeeping genes¹⁸ and genes identified by antibody screening^{15,16}. Regions of low

Table 2 Summary of predicted features on *P. falciparum* chromosome 3 and comparison with chromosome 2 (ref. 7)

	Chromosome 3	Chromosome 2
Length (kb)	1,060	945
No. of predicted proteins	215	209
No. of tRNA genes	2	1
Gene density (kb per gene)	4.8	4.5
Predicted genes with introns (%)	47.4	43.1
Percentage (A + T)		
Overall	80.0	80.3
Exons	76.8	75.7
Introns	84.6	86.7
Predicted protein features		
Signal peptides	50 (23%)	47 (22%)
At least one transmembrane domain	86 (40%)	90 (43%)
Multiple transmembrane domains	48 (22%)	27 (13%)
Coiled-coil domains	106 (49%)	111 (53%)
Low-complexity regions	202 (94%)	155 (74%)
Completely low complexity	0 (0%)	17 (8%)
Detectable homologues in other species	71 (33%)	87 (42%)

Genes on chromosome 3 have been predicted using different algorithms from those used for chromosome 2, and analysis of predicted proteins has been performed using different programs and parameters in some cases.

complexity can be divided into two distinct classes: tandem arrays of repeated peptide motifs (such as CSP and CTRP) and homopolymer runs of a single amino-acid residue (such as RNA polymerase II largest subunit and asparagine synthetase). The homopolymer runs represent expansions of amino acids with A/T-rich codons, encoding asparagine, lysine and glutamic acid, and are the predominant type of peptide polymorphism between parasite isolates.

Before the genome project our understanding of gene organization in *P. falciparum* was limited, as few transcriptional units had been characterized^{22,23}. Cloning and sequencing methodologies tended to encourage identification of genes that were highly immunogenic. Most of these genes had a single exon, whereas the remainder had small additional 5' or 3' exons. Analysis of the chromosome 2 and 3 sequences indicates that splicing is a much more common phenomenon in *P. falciparum* than originally thought. Nearly half of the genes on chromosome 3 are predicted to contain at least one intron (102; 47.4%). For 31.9% of these, intron splicing has been confirmed by similarity data to expressed sequence tag (EST) clones, comparative analysis with orthologous genes or by directly using polymerase chain reaction with reverse transcription (RT-PCR). For the majority of spliced genes on chromosome 3 (63 genes), splicing is predicted to be confined to the 2-exon model. Half of the 2-exon gene predictions have a 5' exon length of <100 bp, making the accurate prediction of their initiation ATG codons particularly difficult.

Several predicted genes consist of multiple small exons (up to 15) exhibiting a gene structure more similar to that of higher eukaryotes. We were able to identify five genes of this type on chromosome 3 owing to their similarity to genes in *P. falciparum* or other organisms. PFC0495W is similar to *Eimeria tenella* aspartyl protease; PFC0935C, *N*-acetylglucosamine-1-phosphate transferase, is most similar to its mouse homologue; and PFC0410W is most similar to the YT51 protein expressed in rat brain. The remaining two genes in this category, PFC0110W and PFC0120W, are members of a *P. falciparum* multigene family that encodes proteins implicated in cytoadherence²⁴. These cytoadherence-linked asexual gene (*clag*) paralogues (*clag* 3.2 and *clag* 3.1, respectively) appear to have arisen from gene duplication; other single copies of *clag* have been localized to chromosomes 2, 4 and 9 (ref. 24). Deletion of the *clag* gene on chromosome 9 abolishes cytoadherence, indicating either that the *clag* genes on chromosome 3 are not functionally equivalent to it, or that their transcription is subject to higher order regulation²⁴. Expression of PFC0410W, PFC0495W, PFC0935C

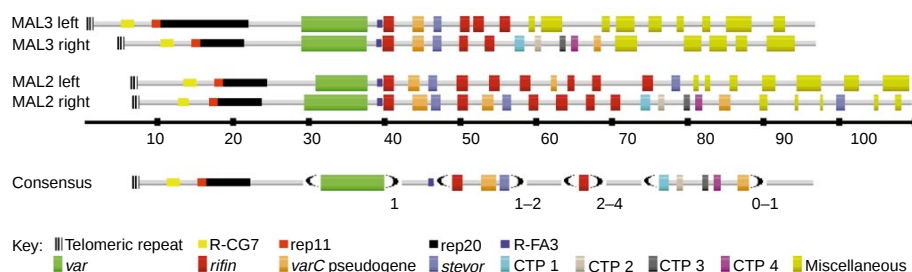


Figure 3 Telomere organization in *P. falciparum*. **a**, The telomeres of chromosomes 2 and 3, showing the higher-order organization of repetitive DNA sequences and telomere-associated multigene families. Four proteins of unknown function encoded on chromosome 3 are similar to proteins encoded on chromosome 2 and have conserved order and orientation (conserved

telomere-encoded proteins: CTPs). These are shown in light blue (PFC1090W/PFB0995W), beige (PFC1085C/PFB990C), grey (PFC1080C/PFB0985C) and fuchsia (PFC1075W/PFB0980W). **b**, A consensus for the arrangement of repetitive DNA sequences and multigene families on *P. falciparum* telomeres based on the telomeres of chromosomes 2 and 3.

and PFC0120W in the asexual blood stages has been confirmed by RT-PCR, and their predicted splice sites corrected by sequencing the cloned RT-PCR products²⁴ (Fig. 2, and data not shown). The complex nature of splicing in this type of gene, and their short exons, makes them difficult to predict in the absence of similarity data; they are predicted inefficiently using the hexamer program (R. Durbin, unpublished software. Documentation, code and data are available from anonymous ftp servers at ftp.sanger.ac.uk/pub/hexamer.; Fig. 2). It is likely that this class of *P. falciparum* genes remains significantly underpredicted from genomic sequence. Sequence data from the RT-PCR products is being used to re-train the current gene-prediction software, but generation of more *P. falciparum* EST sequences or full-length complementary DNA libraries would also facilitate identification of these genes.

We compared the gene-prediction statistics for chromosomes 2 (ref. 7) and 3 (Table 2). As expected, the chromosomes are similar in many respects, including their overall (A + T) composition and coding density. However, the number of genes predicted as having introns is higher for chromosome 3. RT-PCR experiments for some chromosome-3 genes (results not shown) indicate that splicing in *P. falciparum* is actually under-predicted by our current gene-finding methods. Because the initial analysis of chromosome 2 predicts even fewer splicing events⁷, it is likely that splicing has also been underpredicted from the chromosome 2 sequence. It is essential that gene predictions are experimentally confirmed by RT-PCR to generate a larger training set with which to improve gene-finding algorithms.

Direct comparison of the proteins predicted on chromosomes 2 and 3 is hampered because of the different methods used in generating those predictions and the different programs and parameters used in their analysis. For example, the differences observed between these two chromosomes, when comparing predictions of low-complexity sequence, are due at least in part to a different definition of low complexity being used to define the parameters for analysis. Improvement and standardization of gene-finding and protein-analysis methodologies, with subsequent re-analysis of the data from chromosomes 2 and 3, will allow a more accurate comparison of protein features.

Protein targeting

In addition to its nuclear genome, *P. falciparum* contains two organellar genomes, thought to have been acquired as a result of multiple endosymbiotic events. The mitochondrial genome is a 5.9-kb linear molecule present as multiple tandem repeats. A second organellar genome, a 35-kb circular DNA molecule, is located within the apicoplast²⁵. This is an organelle of plastid origin, thought to be unique to apicomplexan parasites, which apparently provides essential metabolic functions²⁶. Gradual loss of genes from the organellar genomes has occurred, such that maintenance of both organelles requires many nuclear-encoded proteins targeted to that

organelle using amino-terminal sequences²⁷. Three predicted proteins on chromosome 3 have N-terminal sequences and implied biological function indicating mitochondrial import (PFC0170C, PFC0225C and PFC0275W). Two predicted proteins have pre-sequences that indicate targeting to the apicoplast (PFC0050C and PFC0310C). An additional three proteins are likely to be localized to an organelle, a conclusion based on their N-terminal sequences and predicted function; however, their precise location cannot be determined from their signal peptides (PFC0470W, PFC495W and PFC725C).

P. falciparum also directs several proteins to the cytoplasm, cytoskeleton and plasma membrane of the infected red blood cell. The proteins implicated as receptors involved in cytoadherence fall into this category; however, the sequences responsible for targeting these molecules remain unidentified.

Telomere structure

Non-coding sequences. The chromosome sequence contains 39 copies of the telomeric repeat sequence (T(G/A)ACCC) at the left telomere and 85 copies at the right telomere, but no attempt has been made to estimate the exact number of copies of this repeat in the intact chromosome. Other repeat sequences (R-CG7, rep11, rep20) occur between the telomere and the gene most proximal to the telomere (var). The rep11 sequence is a new repeat family consisting of an 11-bp tandem repeat located immediately telomeric to the rep20 sequences. The R-FA3 repeat sequence²⁸ maps between the var gene and the adjacent rif gene.

Coding sequences. Of the known multigene families, there are two var genes^{29,30}, one at each telomere, four members of the rif gene family on the left and three on the right chromosome arm and one copy of the stevor family at each telomere. Two members of the clag family²⁴ occur in the left subtelomeric region, separated by a region with similarity to var, possibly representing the site of a previous recombination event. In addition, several pseudogenes occur at both telomeres that have homology to the var 3' exon (the varC genes)³¹. The preservation of these pseudogene sequences in the apparently rapidly evolving subtelomeric regions indicates that they may be biologically significant.

The most telomere-proximal sections of all four *P. falciparum* telomeres sequenced to date show a conserved order of repetitive DNA sequences and multigene family members (Fig. 3). In addition, the right telomere of chromosome 3 shows an extended region of similarity with the right telomere of chromosome 2 (ref. 7; Fig. 3). As well as members of known multigene families, there are several predicted genes that share similarity located on each of these chromosomes, which may represent new telomere-associated multigene families (PFC1075W–PFC1090W and PFB0980W–PFB0995W). The cellular location of proteins encoded by the members of these new multigene families is unknown, although we predict that all eight proteins have N-terminal signal peptides.

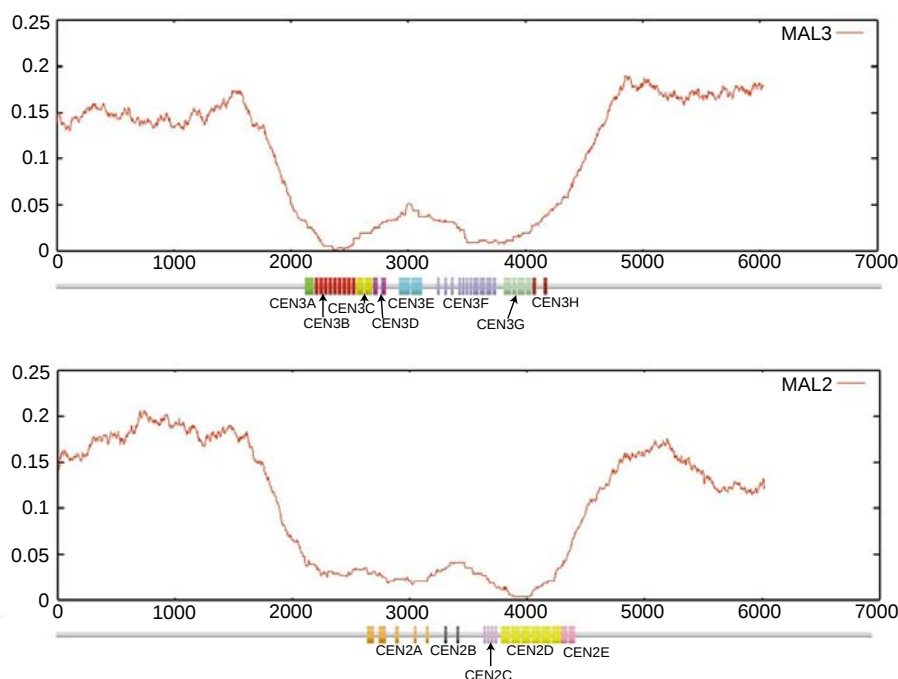


Figure 4 Putative *P. falciparum* centromeres. Plots of (A + T) composition for chromosomes 3 and 2. The Y axes are graduated in units of percentage (G + C) content, ranging from 0% (G + C) to 25% (G + C); the X axis represents DNA sequence in bp. **a**, (A + T) composition of chromosome 3 highlights a region that is extremely (A + T)-rich. Families of repetitive sequences in the predicted

centromere of chromosome 3 are shown as coloured boxes. **b**, A similar region is present on chromosome 2, containing different families of tandem repeat sequences. Alignments of the repeat families can be found at ftp://www.sanger.ac.uk/Projects/P_falciparum/Centromeres/.

Both *rif* and *var* are expressed on the surface of infected red blood cells^{32,33} and undergo antigenic variation, probably as a means of immune evasion³⁴. The products of the *var* locus mediate cytoadherence to a variety of cellular receptors, and thus are important virulence factors. All three of the previously identified multigene families show extreme sequence polymorphism. The clustering of genes sharing similar cellular location and control is likely to be of mechanistic significance and indicates that the newly identified members of the cluster should be studied further. In addition, the regions of shared similarity between the telomeres of chromosomes 2 and 3 indicate that recombination within these regions may be frequent, and could partly explain the extensive polymorphism seen in members of the *stevor*, *rif* and *var* families.

Centromere

No *P. falciparum* centromere has been identified. Centromeres in other eukaryotic species^{35,36} can be divided into two distinct classes: the point centromere seen in budding yeasts, such as *S. cerevisiae*, and the regional centromeres present in *Schizosaccharomyces pombe* and higher eukaryotes. The minimal functional unit of the *S. cerevisiae* centromere consists of an ~125-bp region comprising three parts, the outer CDEI and CDEIII domains that flank a central (A + T)-rich region. *S. cerevisiae* centromeres are located in chromosomal regions of relatively low gene density and high (A + T) composition. The regional class of centromere has been most comprehensively characterized in the fission yeast *S. pombe*. Regional centromeres have characteristic complex arrays of repetitive sequences occurring over an extended region. The three centromeres of *S. pombe* span regions of 40–100 kb, each having a central (A + T)-rich core flanked by chromosome-specific repeat sequences^{37,38}. Sequence in regional centromeres is very variable and this has hampered attempts to find functional centromere sequences in higher eukaryotes³⁵.

We have identified a region on chromosome 3 that is currently the best candidate for the chromosome centromere. This region is

extremely (A + T)-rich, 97.3% over 2.6 kb, with a central region having a slightly higher (G + C) content (Fig. 4a). Analysis of *P. falciparum* chromosome 2 reveals a region with similar structure (Fig. 4b). Both regions have very low coding potential; the nearest predicted genes are 12 kb apart on chromosome 3 (PFC0610c and PFC0615w) and 9 kb apart on chromosome 2 (ref. 7) (PFB0490c and PFB0495w). If these regions are chromosome centromeres, they have a structure more characteristic of regional rather than point centromeres. However, they would represent very compact regional centromeres, with an extremely short core sequence. Analysis of this region on both chromosomes highlights several chromosome-specific tandem repeats (Fig. 4).

Early release of sequence data

Even before this chromosome was completed, several groups had demonstrated the utility of timely release of sequence data in unfinished form. During the course of the project, chromosome 3 sequences were used to confirm chromosome location and sequence data generated independently¹⁹, to identify new members of *P. falciparum* gene families³⁹ and to identify members of families of genes conserved in other genera⁴⁰. □

Methods

Sequencing. *P. falciparum* DNA is denatured at relatively low temperatures because of its extreme (A + T) content⁵, so we had to adapt many standard protocols for this project. DNA used in library preparation was not exposed to either ethidium bromide or ultraviolet (UV) transillumination and we minimized the temperatures to which the DNA was exposed. Gel slices containing chromosome 3 were excised from pulsed-field gels and DNA was extracted from low-melting-point agarose using a modified agarase protocol. After equilibrating with TE buffer, pH 7.4, for several hours, gel slices were loaded into a 2-ml syringe and forced through a 26G needle. Agarase buffer was added to a final concentration of 1×, and 8U β-agarase was added per ml of sample. After incubating at 37 °C for 3 h, the sample was extracted with TE-buffered phenol and the DNA recovered by ethanol precipitation. Libraries were

prepared by fragmenting the DNA by sonication and cloning into pUC18. Sequences were generated using pUC clones with both forward and reverse primers using dye-terminator chemistry.

Assembly. Because of the size and complexity of the project, we modified the standard strategy used for sequence assembly. Initially, sequences generated by the WCS and the YAC skims were assembled using the Phrap assembly program (P. Green, unpublished software), which had been adjusted to handle the large number of reads generated. We divided the chromosome into eight sections based on the YAC map, each section having a separate working database. Contigs that did not contain YAC reads were directed to a repository database, from which they could be recovered once their location had been identified. We expected to have a large number of single reads in the assembly that originated from other *P. falciparum* chromosomes, as the library generated for chromosome 3 was estimated as being 87% pure, based on the hybridization of shotgun reads to chromosome blots. Thus, single reads were not incorporated into any of the databases, but again they could be recovered if necessary. In collaboration with the other laboratories contributing to the Malaria Genome Sequencing Project, the reads generated during the chromosome 3 project that originate from other *P. falciparum* chromosomes will be incorporated into their respective chromosomes as the sequencing project progresses.

Closure. Gaps in the initial assembly were filled by several methods. Initially, oligonucleotide walking from pUC clones bridging contigs was used to fill many of the gaps. A second approach was to perform combinatorial PCR between contigs mapped to the same chromosome region by YAC-derived reads. Products generated by combinatorial PCR were sequenced by oligonucleotide walking. For gaps that could not be filled by PCR, further pUC clones proximal to the gaps were selected by hybridization of a gridded 15x chromosome 3 pUC library to radiolabelled oligonucleotides selected at contig ends⁴¹. Regions of extreme (A + T) composition were resolved by either generating transposon libraries or cloning as very small fragments (50–500 bp) into m13mp18.

Analysis. Completed sections of chromosome 3 were subjected to a series of automatic analyses to reveal possible protein-coding (R. Durbin, unpublished software; P. Green, & L. Hillier, unpublished software) and tRNA⁴² genes, similarities to ESTs⁴³, other proteins^{44,45} and repeat/multigene families. The results were collated in a genome database (ACeDB) that merges overlapping sequences to provide a single contiguous view of the entire chromosome. Documentation, code and data for ACeDB are available from anonymous ftp servers at ftp.sanger.ac.uk/pub/acedb and ncbi.nlm.nih.gov/repository/acedb. Data from the various analyses were viewed interactively through the ACeDB annotator's graphical workbench. GENEFINDER (P. Green & L. Hillier, unpublished software) predictions were confirmed or adjusted to incorporate protein, cDNA and EST matches, hexamer coding potential and repetitive sequences (see http://www.sanger.ac.uk/Projects/P_falciparum/Methods_Analysis for full details of analysis protocols and parameters). We utilized the protein family databases Pfam²⁰ to classify common protein domains in the malaria genome. A number of web-based tools were used to identify transmembrane regions, coiled coil domains, signal peptides and overall suggestions of protein localization (http://www.sanger.ac.uk/Projects/P_falciparum/Toolkit).

Data release. Sequence data generated by the chromosome 3 project were released continuously and were available for searching using the on-site BLAST server and downloading by ftp without restriction. The fully annotated sequence is available for browsing and downloading from http://www.sanger.ac.uk/Projects/P_falciparum. Unfinished sequence data from the remaining eight *P. falciparum* chromosomes in progress at the Sanger Centre can also be accessed through this site.

Received 28 May; accepted 6 July 1999.

1. World Health Organisation *A Global Strategy for Malaria Control* (World Health Organisation, Geneva, 1993).
2. MacPherson, G. G., Warrell, M. J., White, N. J., Looareesuwan, S. & Warrell, D. A. Human cerebral malaria. A quantitative ultrastructural analysis of parasitised erythrocyte sequestration. *Am. J. Pathol.* **119**, 385–401 (1985).
3. White, N. J. The treatment of malaria. *N. Engl. J. Med.* **335**, 800–806 (1996).
4. Hoffman, S. L. et al. Funding for malaria genome sequencing. *Nature* **387**, 647 (1997).
5. Pollack, Y., Katzen, A. L., Spira, D. T. & Golenser, J. The genome of *Plasmodium falciparum*. I: DNA base composition. *Nucleic Acids Res.* **10**, 539–546 (1982).
6. Churcher, C. et al. The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IX. *Nature* (suppl.) **387**, 84–87 (1997).
7. Gardner, M. J. et al. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132 (1998).

8. Thompson, J. K. & Cowman, A. F. AYAC contig and high resolution restriction map of chromosome 3 from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **90**, 537–542 (1997).
9. Triglia, T. & Kemp, D. J. Large fragments of *Plasmodium falciparum* DNA can be stable when cloned in yeast artificial chromosomes. *Mol. Biochem. Parasitol.* **44**, 207–212 (1991).
10. Gardiner, K. et al. YAC analysis and minimal tiling path construction for chromosome 21q. *Somat. Cell. Mol. Genet.* **21**, 399–414 (1995).
11. Vaudin, M. et al. The construction and analysis of M13 libraries prepared from YAC DNA. *Nucleic Acids Res.* **23**, 670–674 (1995).
12. Walliker, D. et al. Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236**, 1661–1666 (1987).
13. Aston, C., Mishra, B. & Schwartz, D. C. Optical mapping and its potential for large scale sequencing projects. *Trends Biotechnol.* **17**, 297–302 (1999).
14. The *C. elegans* sequencing consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
15. Dame, J. B. et al. The structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. *Science* **225**, 593–599 (1984).
16. Enea, V. et al. DNA cloning of *Plasmodium falciparum* circumsporozoite gene: amino acid sequence of the repetitive epitope. *Science* **225**, 628–630 (1984).
17. Trotter, E., Triglia, T. & Cowman, A. F. Molecular cloning of a gene from *Plasmodium falciparum* that codes for a protein sharing motifs found in adhesive molecules from mammals and *Plasmodia*. *Mol. Biochem. Parasitol.* **74**, 129–141 (1995).
18. Li, W. B. et al. An enlarged largest subunit of *Plasmodium falciparum* RNA polymerase II defines conserved and variable RNA polymerase domains. *Nucleic Acids Res.* **17**, 9621–9636 (1989).
19. Berriman, M. & Fairlamb, A. H. Detailed characterisation of a cyclophilin from the human malaria parasite *Plasmodium falciparum*. *Biochem. J.* **334**, 437–445 (1998).
20. Bateman, A. et al. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res.* **27**, 260–262 (1999).
21. Wootton, J. C. & Federhen, S. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.* **266**, 554–571 (1996).
22. Lanzer, M., de Bruijn, D. & Ravetch, J. V. Transcriptional mapping of a 100 kb locus of *Plasmodium falciparum* identifies a region in which transcription terminates and reinitiates. *EMBO J.* **11**, 1949–1955 (1992).
23. Horrocks, P., Dechering, K. & Lanzer, M. Control of gene expression in *Plasmodium falciparum*. *Mol. Biol. Parasitol.* **95**, 171–181 (1998).
24. Holt, D. C. et al. The *clag* gene family of *Plasmodium falciparum*: are there roles other than cytoadherence? *Int. J. Parasitol.* (in the press).
25. Wilson, R. J. M. et al. Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* **261**, 155–172 (1996).
26. Soldati, D. The apicoplast as a potential therapeutic target in *Toxoplasma* and other apicomplexan parasites. *Parasitol. Today* **15**, 5–7 (1999).
27. Waller, R. F. et al. Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **95**, 12352–12357 (1998).
28. De Bruijn, D., Lanzer, M. & Ravetch, J. V. The polymorphic subtelomeric regions of *Plasmodium falciparum* chromosomes contain arrays of repetitive sequence elements. *Proc. Natl Acad. Sci. USA* **91**, 619–623 (1994).
29. Thompson, J. K. et al. The chromosomal organisation of the *Plasmodium falciparum* var gene is conserved. *Mol. Biochem. Parasitol.* **87**, 49–60 (1997).
30. Fischer, K. et al. Expression of var genes located within polymorphic subtelomeric domains of *Plasmodium falciparum* chromosomes. *Mol. Cell. Biol.* **17**, 3679–3686 (1997).
31. Carcy, B. et al. A large multigene family expressed during the erythrocytic schizogony of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **68**, 221–233 (1997).
32. Leech, J. H., Barnwell, J. W., Miller, L. H. & Howard, R. J. Identification of a strain-specific malarial antigen exposed on the surface of *Plasmodium falciparum*-infected erythrocytes. *J. Exp. Med.* **159**, 1567–1575 (1984).
33. Kyes, S., Rowe, A., Kriek, N. & Newbold, C. I. RiFins: A second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* (in the press).
34. Newbold, C. I. et al. PfEMP1, polymorphism and pathogenesis. *Ann. Trop. Med. Parasitol.* **91**, 551–557 (1997).
35. Pluta, A. F., Mackay, A. M., Ainsztein, A. M., Goldberg, I. G. & Earnshaw, W. C. The centromere: hub of chromosomal activities. *Science* **270**, 1591–1594 (1995).
36. Clarke, L. Centromeres: proteins, protein complexes, and repeated domains at centromeres of simple eukaryotes. *Curr. Opin. Genet. Dev.* **8**, 212–218 (1998).
37. Baum, M., Ngan, V. K. & Clarke, L. The centromeric K-type repeat and the central core are together sufficient to establish a functional *Schizosaccharomyces pombe* centromere. *Mol. Biol. Cell* **5**, 747–761 (1994).
38. Takahashi, K. et al. A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. *Mol. Biol. Cell* **3**, 819–835 (1992).
39. Cheng, Q. et al. *stevor* and *rif* are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. *Mol. Biochem. Parasitol.* **97**, 161–176 (1998).
40. Garcia, A., Cayla, X., Barik, S. & Langsley, G. A family of PP2 phosphatases in *Plasmodium falciparum* and parasitic protozoa. *Parasitol. Today* **15**, 90–91 (1999).
41. McMurray, A. A., Sulston, J. E. & Quail, M. A. Short-insert libraries as a method for problem solving in genome sequencing. *Genome Res.* **8**, 562–566 (1998).
42. Lowe, T. M. & Eddy, S. R. tRNA scan-SE: a program for improved detection of transfer RNA genes in genome sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
43. Mott, R. EST-GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. *Comput. Appl. Biosci.* **13**, 477–478 (1997).
44. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **15**, 403–410 (1990).
45. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

Acknowledgements. We thank the staff in the subcloning, gel-pouring and media groups for their support; the computer support and software development groups; J. Thompson and A. Cowman for gifts of YAC clones and for advice; D. Schwartz for optical and X. Su for genetic map information; B. Streipen, D. Holt and D. Kemp for communicating results before publication; Z. Christodoulou, R. Pinches and S. Lee for technical assistance; the other members of the Malaria Genome Sequencing consortium for useful discussions; K. Rutherford and R. Summers for help with Fig. 1; and J. Parkhill for critical reading of the manuscript. This work was funded by the Wellcome Trust.

Correspondence and requests for materials should be addressed to S.B. (e-mail: sharen@sanger.ac.uk). Sequence accession numbers and other information can be found on http://www.sanger.ac.uk/Projects/P_falciparum.

A geometric distance to the galaxy NGC4258 from orbital motions in a nuclear gas disk

J. R. Herrnstein^{*†}, J. M. Moran[†], L. J. Greenhill[†],
P. J. Diamond^{*‡}, M. Inoue[§], N. Nakai[§], M. Miyoshi^{||},
C. Henkel[¶] & A. Riess[#]

^{*} National Radio Astronomy Observatory, PO Box O, Socorro, New Mexico 87801, USA

[†] Harvard-Smithsonian Center for Astrophysics, Mail Stop 42, 60 Garden Street, Cambridge, Massachusetts 02138, USA

[‡] Merlin and VLBI National Facility, Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK

[§] Nobeyama Radio Observatory, National Astronomical Observatory, Minamimaki, Minamisaku, Nagano 384-13, Japan

^{||} VERA Project Office, National Astronomical Observatory, Mitaka, Tokyo, 181-8588, Japan

[¶] Max Planck Institut für Radioastronomie, Auf dem Hugel 69, D-53121, Bonn, Germany

[#] Department of Astronomy, University of California at Berkeley, Berkeley, California 94720, USA

The accurate measurement of extragalactic distances is a central challenge of modern astronomy, being required for any realistic description of the age, geometry and fate of the Universe. The measurement of relative extragalactic distances has become fairly

routine, but estimates of absolute distances are rare¹. In the vicinity of the Sun, direct geometric techniques for obtaining absolute distances, such as orbital parallax, are feasible, but such techniques have hitherto been difficult to apply to other galaxies. As a result, uncertainties in the expansion rate and age of the Universe are dominated by uncertainties in the absolute calibration of the extragalactic distance ladder². Here we report a geometric distance to the galaxy NGC4258, which we infer from the direct measurement of orbital motions in a disk of gas surrounding the nucleus of this galaxy. The distance so determined— 7.2 ± 0.3 Mpc—is the most precise absolute extragalactic distance yet measured, and is likely to play an important role in future distance-scale calibrations.

NGC4258 is one of 22 nearby active galactic nuclei (AGN) known to possess nuclear water masers (the microwave equivalent of lasers). The enormous surface brightnesses ($\geq 10^{12}$ K), relatively small sizes ($\leq 10^{14}$ cm) and narrow linewidths (a few km s^{-1}) of these masers make them ideal probes of the structure and dynamics of the molecular gas in which they reside. Very-long-baseline interferometry (VLBI) observations of the NGC4258 maser have provided the first direct images of an AGN accretion disk, revealing a thin, subparsec-scale, differentially rotating warped disk in the nucleus of this relatively weak Seyfert 2 AGN^{3–6}. Two distinct populations of masers exist in NGC4258. The first are the high-velocity masers. These masers amplify their own spontaneous emission and are offset $\pm 1,000 \text{ km s}^{-1}$ and $4.7\text{--}8.0 \text{ mas}$ ($0.16\text{--}0.28 \text{ pc}$ for a distance of 7.2 Mpc) on either side of the disk centre. The keplerian rotation curve traced by these masers requires a central binding mass (M), presumably in the form of a supermassive black hole, of $(3.9 \pm 0.1) \times 10^7 (D/7.2 \text{ Mpc})(\sin i_s/\sin 82^\circ)^{-2}$

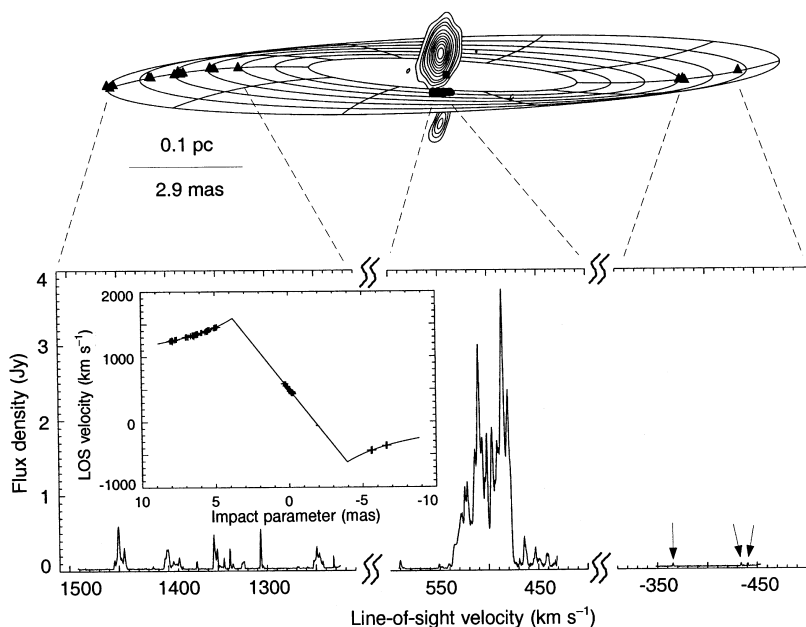


Figure 1 The NGC4258 water maser. The upper panel shows the best-fitting warped-disk model superposed on actual maser positions as measured by the VLBA of the NRAO, with top as North. The filled square marks the centre of the disk, as determined from a global disk-fitting analysis⁸. The filled triangles show the positions of the high-velocity masers, so called because they occur at frequencies corresponding to Doppler shifts of $\sim \pm 1,000 \text{ km s}^{-1}$ with respect to the galaxy systemic velocity of $\sim 470 \text{ km s}^{-1}$. This is apparent in the VLBA total power spectrum (lower panel). The inset shows line-of-sight (LOS) velocity versus impact parameter for the best-fitting keplerian disk, with the maser data superposed. The high-velocity masers trace a keplerian curve to better than 1%. Monitoring of these features indicates that they drift by less than $\sim 1 \text{ km s}^{-1} \text{ yr}^{-1}$ (refs 14–16) and requires that they lie within $5\text{--}10^\circ$ of the midline, the intersection of the disk with the plane of the sky. The LOS velocities of the systemic masers are

centred about the systemic velocity of the galaxy. The positions (filled circles in upper panel) and LOS velocities of these masers imply that they subtend $\sim 8^\circ$ of disk azimuth centred about the LOS to the central mass, and the observed acceleration ($8\text{--}10 \text{ km s}^{-1} \text{ yr}^{-1}$) of these features^{14,15} unambiguously places them along the near edge of the disk. The approximately linear relationship between systemic maser impact parameter and LOS velocity demonstrates that the disk is very thin¹⁷ (aspect ratio $\leq 0.2\%$) and that these masers are confined to a narrow annulus in the disk. The magnitude of the velocity gradient (Ω_s) implies a mean systemic radius, r_s , of 3.9 mas which, together with the positions of the high-velocity masers, constrains the disk inclination, i_s , to be $\sim 82 \pm 1^\circ$ (90° for edge-on). Finally, VLBA continuum images^{7,13} are included as contours in the upper panel. The 22-GHz radio emission traces a sub-parsec-scale jet elongated along the rotation axis of the disk and well-aligned with a luminous, kiloparsec-scale jet¹⁸.

solar masses ($M_\odot$) where D is the distance to NGC4258 and i_s is the disk inclination. Because the high-velocity masers lie in the plane of the sky, they should to first order remain stationary as the disk rotates. The second population of masers are the systemic masers. These masers are positioned along the near edge of the disk and amplify the background jet emission evident in Fig. 1 (ref. 7). A fundamental and as-yet untested prediction of the maser disk model is that the systemic masers should drift with respect to a fixed point on the sky by a few $10 \mu\text{as yr}^{-1}$ as the disk rotates at $\sim 1,000 \text{ km s}^{-1}$.

NGC4258 was observed at 4–8-month intervals between 1994 and 1997 with the Very Long Baseline Array (VLBA) of the NRAO in order to search for the expected motions. As the maser emission is essentially continuous across the envelope of systemic maser emission, we are forced to rely on structure in the systemic spectrum to isolate and track individual maser features. The assumption that distinct peaks in the spectrum correspond to individual clumps of gas is justified by the successful tracking of maser accelerations in single-dish monitoring programmes, and leads to an identification of 20–35 potentially trackable systemic masers in each epoch. These features are too densely packed in position and velocity to be individually tracked with any reliability. However, the resolution and sampling are sufficient to detect bulk rotation in the system, and we have developed a bayesian pattern-matching analysis tool to track inter-epoch shifts in the positions and velocities of the systemic masers, as a whole⁸. The pattern-matching analysis assumes that the systemic masers are randomly and narrowly scattered about an average radius, $\langle r_s \rangle$, of 3.9 mas, as indicated by the global disk-fitting analysis (see Fig. 1). The precise magnitude of the radial scatter is set so as to maximize the overall likelihood of the tracking analysis. To evaluate the likelihood of a given bulk proper motion ($\langle \dot{\theta}_x \rangle$) or acceleration ($\langle \dot{v}_{\text{LOS}} \rangle$), the pattern-matching procedure must compute the likelihood that each individual maser has in fact moved by $\langle \dot{\theta}_x \rangle$ or $\langle \dot{v}_{\text{LOS}} \rangle$. This leads to robust estimates for the ‘trackability’ of each maser. Figure 2 shows the best-fitting acceleration and proper-motion tracks for the most reliably trackable systemic masers. Figure 3 shows the overall probability density functions (PDFs) for $\langle \dot{v}_{\text{LOS}} \rangle$ and $\langle \dot{\theta}_x \rangle$. The PDFs indicate a highly significant detection of bulk motion in the disk, and from them we conclude that $\langle \dot{v}_{\text{LOS}} \rangle = 9.3 \pm 0.3 \text{ km s}^{-1} \text{ yr}^{-1}$ and $\langle \dot{\theta}_x \rangle = 31.5 \pm 1 \mu\text{as yr}^{-1}$, where these (and all subsequent) uncertainties are 1σ values. The latter result is consistent with expectations, and is the first detection of transverse motion in the NGC4258 accretion disk. We note that the pattern-matching algorithm has been verified on a number of simulated data sets with feature densities, and spectral and spatial resolutions, comparable to those of the true data.

To convert the maser proper motions and accelerations into a geometric distance, we express $\langle \dot{\theta}_x \rangle$ and $\langle \dot{v}_{\text{LOS}} \rangle$ in terms of the distance and four disk parameters:

$$\langle \dot{\theta}_x \rangle = 31.5 \left[\frac{D_6}{7.2} \right]^{-1} \left[\frac{\Omega_s}{282} \right]^{1/3} \left[\frac{M_{7.2}}{3.9} \right]^{1/3} \left[\frac{\sin i_s}{\sin 82.3^\circ} \right]^{-1} \left[\frac{\cos \alpha_s}{\cos 80^\circ} \right] \mu\text{as yr}^{-1} \quad (1)$$

and

$$\langle \dot{v}_{\text{LOS}} \rangle = 9.2 \left[\frac{D_6}{7.2} \right]^{-1} \left[\frac{\Omega_s}{282} \right]^{4/3} \left[\frac{M_{7.2}}{3.9} \right]^{1/3} \left[\frac{\sin i_s}{\sin 82.3^\circ} \right]^{-1} \text{ km s}^{-1} \text{ yr}^{-1} \quad (2)$$

Here D_6 is the distance in Mpc, α_s is the disk position angle (East of North) at $\langle r_s \rangle$, and $M_{7.2}$ is $M/D \sin^2 i_s$ as derived from the high-velocity rotation curve and evaluated at $D = 7.2 \text{ Mpc}$ and $i_s = 82.3^\circ$ (in units of $10^7 M_\odot$). $\Omega_s \equiv (GM_{7.2}/\langle r_s \rangle^3)^{1/2}$ is the projected disk angular velocity at $\langle r_s \rangle$ as determined by the slope of the systemic position–velocity gradient (in units of $\text{km s}^{-1} \text{ mas}^{-1}$; see Fig. 1). In the denominators of each of the terms of equations (1) and (2), we include *a priori* estimates for each of these disk parameters, derived

directly from the positions and velocities of the masers.

When the *a priori* disk parameter estimates are used, the proper motions and accelerations yield independent distance estimates, through equations (1) and (2), of $7.2 \pm 0.2 \text{ Mpc}$ and $7.1 \pm 0.2 \text{ Mpc}$, respectively. The quoted uncertainties are effectively the uncertainties in $\langle \dot{\theta}_x \rangle$ and $\langle \dot{v}_{\text{LOS}} \rangle$ recast in terms of distance, and as such are

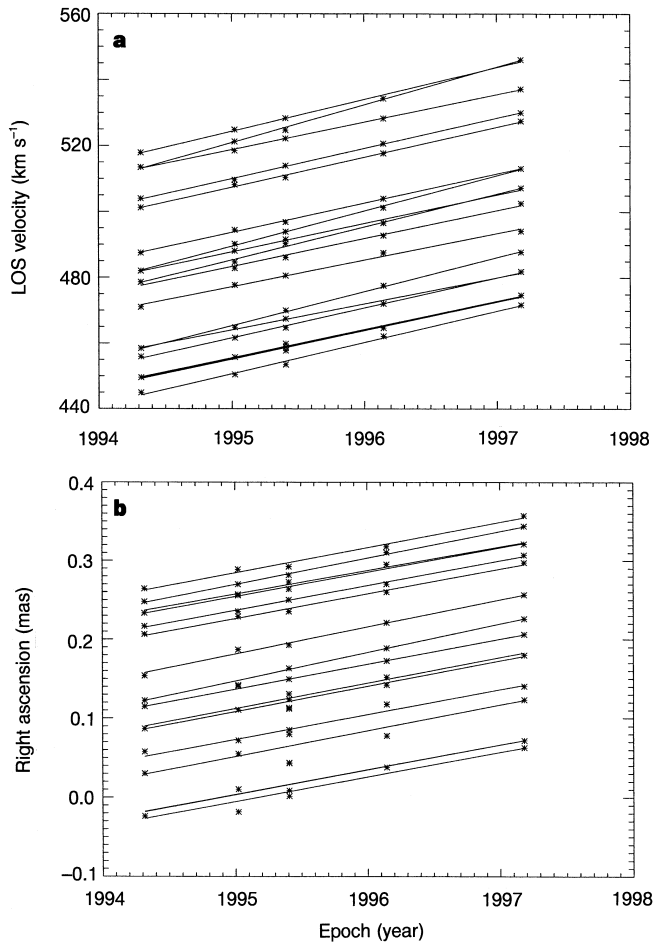


Figure 2 Line-of-sight (LOS) velocities (**a**) and right ascensions (**b**) at the peaks of the systemic maser spectrum. Values of these parameters are shown for each of the five VLBA epochs. Only those features deemed reliably trackable by the pattern-matching analysis are included. All epochs included the Very Large Array, phased to act as a single large aperture. In addition, the final epoch utilized the Effelsberg 100-m telescope. In each epoch, the VLBA correlator provided cross-power spectra with channel spacings of 0.22 km s^{-1} . The NGC4258 masers are characterized by linewidths of $\sim 1\text{--}3 \text{ km s}^{-1}$. All spectra were phase-referenced, via self-calibration, to a strong systemic maser to stabilize the interferometer against atmospheric pathlength fluctuations, and synthesis images were constructed for each spectral channel using conventional restoration techniques¹⁹. Error bars have been omitted for clarity. The uncertainties in individual LOS velocity estimates are dominated by line blending in the spectrum. There is an average scatter of $\sim 0.4 \text{ km s}^{-1}$ about the best-fitting acceleration tracks. The masers are spatially unresolved and relative positions have been measured to a precision of $\sim 0.5\theta_B/\text{SNR}$, where SNR is the signal-to-noise ratio and θ_B represents the $0.6 \times 0.9 \text{ mas}$, approximately North–South, synthesized beam. Relative positional accuracies typically ranged from 0.5 to $10 \mu\text{as}$. All positions are relative to a fixed point along the systemic position–velocity gradient. Our ability to precisely align this structure amongst all epochs suggests that it does indeed remain fixed in time. The best-fitting acceleration and proper-motion tracks (solid lines) indicate average drifts of $9.3 \text{ km s}^{-1} \text{ yr}^{-1}$ and $31.5 \mu\text{as yr}^{-1}$ in velocity and position, respectively. The scatter in the individual proper motions and accelerations about these average values is consistent with 0.2 mas scatter in the radii of the systemic masers about $\langle r_s \rangle$.

purely statistical in nature. The excellent agreement between the proper motion and acceleration distances for *a priori* values of the disk parameters is confirmation of the disk model itself, and establishes the NGC4258 keplerian disk as a fully self-consistent, dynamical model incorporating the positions, line-of-sight (LOS) velocities, proper motions, and accelerations of all the NGC4258 masers. We note that a preliminary geometric distance estimate, based on accelerations alone and a single VLBA epoch, yielded $D = 6.4 \pm 0.9$ Mpc (ref. 4). Hence, the old and new distances are consistent at the 1σ level. The discrepancy between the estimates is explained by the fact that the original distance estimate assumed an average systemic maser radius $\sim 10\%$ larger than the value indicated by the newer data and the more sophisticated disk models.

Uncertainties in the disk parameters contribute to systematic uncertainties in the distance estimate. We derive a composite geometric distance estimate using equations (1) and (2), and from the $\langle \dot{v}_{\text{LOS}} \rangle$ and $\langle \dot{\theta}_s \rangle$ PDFs of Fig. 3 and the *a priori* estimates for the disk parameters and their associated uncertainties. The result is a geometric distance estimate of 7.2 ± 0.3 Mpc, where the quoted uncertainty now incorporates all statistical terms associated with tracking motions in the disk as well as systematics arising from disk parameter uncertainties. An uncertainty of $\sim 5\%$ in Ω_s is the dominant contributor to the latter, producing fractional uncertainties in the acceleration and proper-motion distances of 6.7% and 1.7%, respectively. In total, disk-model systematics contribute an additional 0.26 Mpc (in quadrature) to the distance error budget derived from purely statistical considerations.

The NGC4258 geometric distance is the most precise, absolute extragalactic distance measured to date and, being independent of all other distance indicators, it represents a new calibration point for

the extragalactic distance ladder. This geometric distance is consistent with pre-existing H-band Tully–Fisher (7.1 ± 1.1 Mpc; ref. 9), blue Tully–Fisher (7.9 ± 1.8 Mpc; ref. 10), and luminosity class (8.4 ± 2.2 Mpc; ref. 11) distance estimates, all of which rely on the Cepheid period–luminosity relationship for absolute calibration. Efforts are under way to determine directly a Cepheid distance to NGC4258 from Hubble Space Telescope (HST) observations of the galaxy. The NGC4258 geometric distance is at present the most precise means for directly calibrating distances obtained by HST Cepheid observations.

We emphasize that the above error budget considers statistical and systematic uncertainties within the framework of a thin, circular keplerian disk in which the masers trace the orbital motions of discrete clumps of gas. As always, it is difficult to estimate any additional systematic uncertainties that might exist as a result of imperfections in the model itself. The potential effect of eccentricity on the distance error budget depends on the assumed distribution of accretion-disk eccentricities in AGN in general. Theories of accretion-disk circularization and eccentric orbit stability remain relatively immature. Detailed modelling of the optical emission lines from a large sample of AGN suggests eccentricities within the broad line region of less than ~ 0.5 (ref. 12), and we formulate an eccentricity based on these observations. The eccentricity of the NGC4258 disk is further constrained by the symmetry of the maser emission about the disk centre in both position and velocity¹³. These additional constraints lead to an expected eccentricity of zero with a probable error of 0.1, a negligible bias in the distance estimate, and a systematic uncertainty in the distance of 0.4 Mpc. Hence our distance estimate, including this uncertainty in the eccentricity, is 7.2 ± 0.5 Mpc. Finally, we cannot unambiguously rule out contamination of the maser dynamics by some non-kinematical contribution, such as travelling density waves within the disk. However, given the complexity (30 systemic masers across 8° of disk azimuth) and the stability ($\geq 70\%$ of the features persisting) of the pattern we have tracked, orbital motion is certainly the simplest explanation. $\square$

Received 23 February; 17 June 1999.

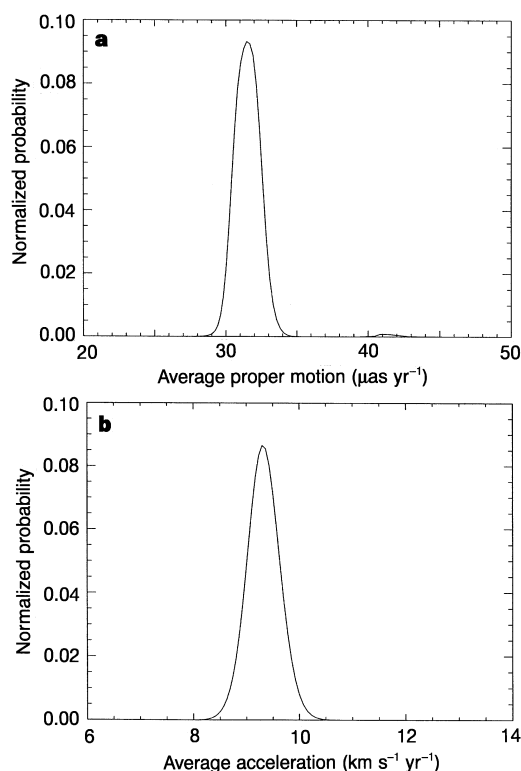


Figure 3 Systemic maser bulk proper motion ($\langle \dot{\theta}_s \rangle$; **a**) and acceleration ($\langle \dot{v}_{\text{LOS}} \rangle$; **b**) probability density functions. These were derived using the bayesian pattern-matching analysis described in the text. The curves were generated using all the maser features in each epoch. The uncertainties in $\langle \dot{\theta}_s \rangle$ and $\langle \dot{v}_{\text{LOS}} \rangle$ as derived from these PDFs include measurement uncertainties in the maser positions and velocities, but they are dominated by ambiguities in tracking specific maser features.

- Jacoby, G. H. *et al.* A critical review of selected techniques for measuring extragalactic distances. *Publ. Astron. Soc. Pacif.* **104**, 599–662 (1992).
- Madore, B. F. *et al.* The Hubble Space Telescope key project on the extragalactic distance scale. XV. A Cepheid distance to the Fornax cluster and its implications. *Astrophys. J.* **515**, 29–41 (1999).
- Watson, W. D. & Wallin, B. K. Evidence from masers for a rapidly rotating disk at the nucleus of NGC4258. *Astrophys. J.* **432**, L35–L38 (1994).
- Miyoshi, M. *et al.* Evidence for a black hole from high rotation velocities in a sub-parsec region of NGC4258. *Nature* **373**, 127–129 (1995).
- Greenhill, L. G. *et al.* Detection of a subparsec diameter disk in the nucleus of NGC4258. *Astrophys. J.* **440**, 619–627 (1995).
- Herrnstein, J. R., Greenhill, L. J. & Moran, J. M. The warp in the subparsec molecular disk in NGC4258 as an explanation for persistent asymmetries in the maser spectrum. *Astrophys. J.* **468**, L17–L20 (1996).
- Herrnstein, J. R. *et al.* Discovery of a subparsec jet 4000 Schwarzschild radii from the central engine of NGC4258. *Astrophys. J.* **475**, L17–L21 (1997).
- Herrnstein, J. R. *Observations of the Sub-Parsec Maser Disk in NGC4258*. Thesis, Harvard Univ. (1997).
- Aaronson, M. *et al.* A catalog of infrared magnitudes and HI velocity widths for nearby galaxies. *Astrophys. J. Suppl.* **50**, 241–262 (1982).
- Richter, O.-G. & Huchtmeier, W. K. Is there a unique relation between absolute (blue) luminosity and total 21 cm linewidth of disk galaxies? *Astron. Astrophys.* **132**, 253–264 (1984).
- Rowan-Robinson, M. *The Cosmological Distance Ladder* (Freeman, New York, 1985).
- Eracleous, M., Livio, M., Halpern, J. P. & Storch-Bergmann, T. Elliptical accretion disks in active galactic nuclei. *Astrophys. J.* **438**, 610–622 (1995).
- Herrnstein, J. R. *et al.* VLBA continuum observations of NGC4258: Constraints on an advection-dominated accretion flow. *Astrophys. J.* **497**, L69–L73 (1998).
- Nakai, N., Inoue, M., Miyaza, K., Miyoshi, M. & Hall, P. Search for extremely high velocity H₂O maser emission in Seyfert galaxies. *Publ. Astron. Soc. Jpn* **47**, 771–799 (1995).
- Greenhill, L. J., Henkel, C., Becker, R., Wilson, T. L. & Wouterloot, J. G. A. Centripetal acceleration within the subparsec nuclear maser disk of NGC4258. *Astron. Astrophys.* **304**, 21–33 (1995).
- Bragg, A. E., Greenhill, L. J., Moran, J. M. & Henkel, C. Acceleration-derived positions of the high-velocity maser features in NGC4258. *Bull. Am. Astron. Soc.* **30**, 1254 (1998).
- Moran, J. M. *et al.* Probing active galactic nuclei with H₂O megamasers. *Proc. Natl Acad. Sci. USA* **92**, 11427–11433 (1995).
- Cecil, G., Wilson, A. S. & DePree, C. Hot shocked gas along the helical jets of NGC4258. *Astrophys. J.* **440**, 181–190 (1995).
- Zensus, J. A., Diamond, P. J. & Napier, P. J. *Very Long Baseline Interferometry and the VLBA* (Astron. Soc. Pacific, San Francisco, 1995).

Acknowledgements. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

Correspondence and requests for materials should be addressed to J.R.H. at NRAO (e-mail: jherrnst@nrao.edu).

Identification of atomic-like electronic states in indium arsenide nanocrystal quantum dots

Uri Banin*, YunWei Cao*, David Katz† & Oded Millo†

* Department of Physical Chemistry and the Farkas Center for Light Induced Processes, † Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel

Semiconductor quantum dots, due to their small size, mark the transition between molecular and solid-state regimes, and are often described as ‘artificial atoms’ (refs 1–3). This analogy originates from the early work on quantum confinement effects in semiconductor nanocrystals, where the electronic wavefunctions are predicted⁴ to exhibit atomic-like symmetries, for example ‘s’ and ‘p’. Spectroscopic studies of quantum dots have demonstrated discrete energy level structures and narrow transition linewidths^{5–9}, but the symmetry of the discrete states could be inferred only indirectly. Here we use cryogenic scanning tunnelling spectroscopy to identify directly atomic-like electronic states with s and p character in a series of indium arsenide nanocrystals. These states are manifest in tunnelling current–voltage measurements as two- and six-fold single-electron-charging multiplets respectively, and they follow an atom-like Aufbau principle of sequential energy level occupation¹⁰.

InAs nanocrystals for this study were prepared using a solution-phase pyrolytic reaction of organometallic precursors. These nanocrystals are nearly spherical in shape, with size controlled between 10 and 40 Å in radius¹¹. The nanocrystal surface is passivated by organic ligands which also provide chemical accessibility for quantum dot (QD) manipulation: for example, to prepare “nanocrystal molecules”^{12,13}, nanocrystal-based light-emitting diodes¹⁴, and to fabricate single-electron tunnelling devices¹⁵. For the tunnelling spectroscopy studies we link the nanocrystals to a gold film via hexane dithiol molecules¹⁶.

Figure 1a (left inset) shows a scanning tunnelling microscope (STM) topographic image of an isolated InAs QD, 32 Å in radius. Also shown in Fig. 1a is a tunnelling current–voltage (*I*–*V*) curve that was acquired after positioning the STM tip above the QD and disabling the scanning and feedback controls, realizing a double-barrier tunnel junction (DBTJ) configuration^{17–19} (Fig. 1a, right inset). A region of suppressed tunnelling current is observed around zero bias, followed by a series of steps at both negative and positive bias. In Fig. 1b we show the tunnelling conductance spectrum (that is, *dI/dV* versus *V*), which is proportional to the tunnelling density of states (DOS)²⁰. A series of discrete single-electron tunnelling peaks is clearly observed, where the separations are determined by both single-electron charging energy (addition spectrum) and the discrete level spacings (excitation spectrum) of the QD. The *I*–*V* characteristics were acquired with the tip retracted from the QD to a distance where the bias voltage is dropped largely across the tip–QD junction. Under these conditions, conduction (valence) band states appear at positive (negative) sample bias, and the excitation peak separations are equal to the real QD level spacings¹⁹.

On the positive-bias side of Fig. 1, two closely spaced peaks are observed immediately after current onset, followed by a larger spacing and a group of six nearly equidistant peaks. We attribute this doublet to tunnelling through the lowest conduction band QD state, where the spacing corresponds to the single-electron charging energy, $E_c = 0.11$ eV. The observed doublet is consistent with the degeneracy of the envelope function of the first conduction band

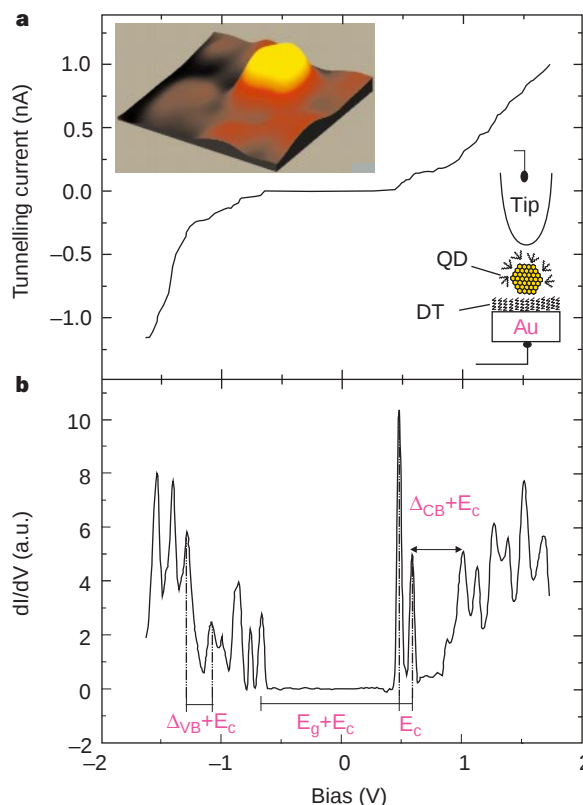


Figure 1 Scanning tunnelling microscopy and spectroscopy of a single InAs nanocrystal 32 Å in radius, acquired at 4.2 K. The nanocrystal quantum dots (QD) are linked to the gold substrate by hexane dithiol molecules (DT), as shown schematically in the right inset. Left inset, a 10 × 10 nm STM topographic image, showing the nanocrystal. For measuring the *I*–*V* characteristics, the STM tip was positioned above the QD, thus realizing a double-barrier tunnel junction configuration. **a**, The tunnelling *I*–*V* characteristic, exhibiting single-electron tunnelling effects. **b**, The tunnelling conductance spectrum, *dI/dV* versus *V*, obtained by numerical differentiation of the *I*–*V* curve (a.u., arbitrary units). The arrows depict the main energy separations: E_c is the single-electron charging energy, E_g is the nanocrystal bandgap, and Δ_{VB} and Δ_{CB} are the spacing between levels in the valence and conduction bands, respectively.

level, $1S_c$, which has s character. This direct relationship between multiplicity of a QD level and the number of addition peaks is substantiated by the observation that the second group consists of six peaks spaced by values close to the observed E_c for the first doublet. The degeneracy of the spherical atomic-like QD levels with envelope-function angular momentum *l* is $2(2l + 1)$, including the electron spin. The second QD conduction band level, $1P_c$, has p character ($l = 1$) and can thus be occupied by six electrons. This sequential level filling resembles the Aufbau principle of building up the lowest-energy electron configuration of an atom¹⁰, directly demonstrating the atomic-like nature of the QD.

The separation between the two groups of peaks, 0.42 eV, is a sum of the level spacing $\Delta_{CB} = 1P_c - 1S_c$, and the charging energy E_c . A value of $\Delta_{CB} = 0.31$ eV is thus obtained. On the negative-bias side, tunnelling through filled dot levels takes place, reflecting the tunnelling DOS of the QD valence band. Again, two groups of peaks are observed. The multiplicity in this case, in contrast with the conduction band, cannot be clearly assigned to specific angular momentum degeneracy. In a manner similar to that described above for the conduction band states, we extract a value of $\Delta_{VB} = 0.10$ eV for the level separation between the two observed valence band states.

In the region of null current around zero bias, the tip and substrate Fermi energies are located within the QD bandgap where the tunnelling DOS is zero. Tunnelling starts when the bias

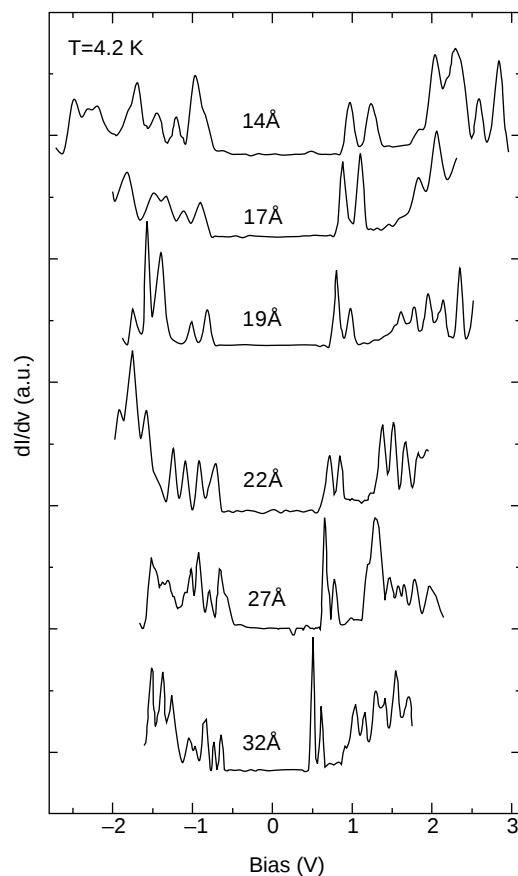


Figure 2 Size evolution of representative tunnelling dI/dV versus V characteristics, displaced vertically. The position of the centre of the observed zero-current gap showed non-systematic variations with respect to zero bias, of the order of 0.2 eV, probably due to variations of local offset potentials. For clarity of presentation, we offset the spectra along the V direction to situate the centres of the observed zero-current gaps at zero bias. The nanocrystal radii are denoted in the figure. The range of displayed voltage for each curve reflects the experimental saturation limit of the detected current.

is large enough to overcome both the bandgap, E_g , and charging energy. E_g is thus extracted by subtracting E_c from the observed spacing between the highest valence band and lowest conduction band peaks, and is equal to 1.02 eV.

The tunnelling conductance spectra for single InAs nanocrystals of several radii are shown in Fig. 2. Two groups of peaks are persistently observed in the positive-bias side (conduction band). The first is always a doublet, consistent with the expected s symmetry of the $1S_c$ level, while the second has higher multiplicity of up to six, consistent with $1P_c$. (Partial indication of such behaviour was also observed previously on electrodeposited CdSe QDs²⁵.) The separation between the two groups, as well as the spacing of peaks within each multiplet, increase with decreasing QD radius. This reflects quantum size effects, on both the nanocrystal energy levels and its charging energy, respectively. On the negative-bias side, two groups of peaks, which exhibit similar quantum confinement effects are also observed. Here we find variations in the group multiplicities between QDs of different size, as well as variations of peak energy spacings within each group. This behaviour is partly due to the fact that the charging energy, E_c , is very close to the level spacing in the valence band Δ_{VB} , as shown in Fig. 3b. In this case, sequential single-electron tunnelling may be either addition to the same valence band state or excitation with no extra charging to the next state. An atomic analogy for this situation can be found in the changing order of electron occupation when

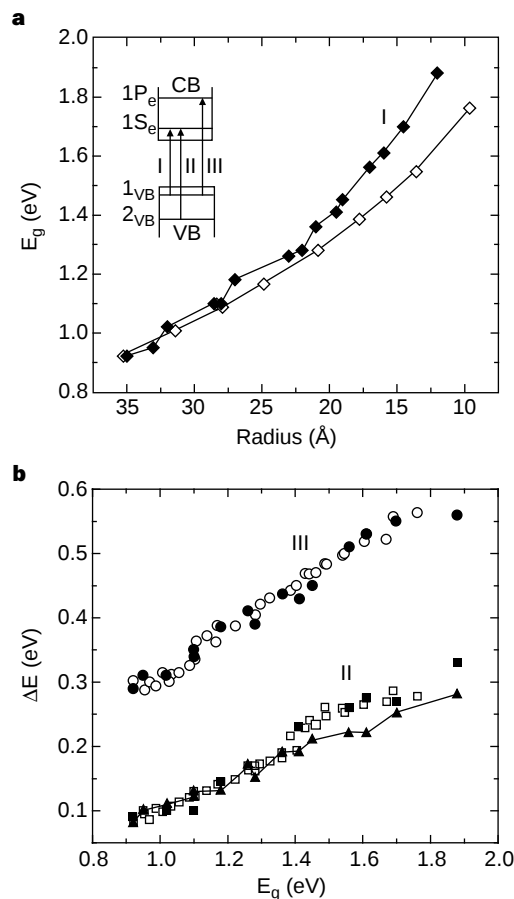


Figure 3 Correlation of optical and tunnelling spectroscopy data for InAs nanocrystals. Inset, diagram of the level structure of the conduction and valence bands, and the relevant strong optical transitions I, II and III. **a**, Comparison of the size dependence of the low-temperature optical bandgap (transition I) from which the excitonic Coulomb interaction was subtracted (open diamonds), with the bandgap measured by the STM (filled diamonds). **b**, Plot of excited transitions versus the bandgap for tunnelling and optical spectroscopy. The two lower data sets (II) depict the correlation between $\Delta_{VB} = 1_{VB} - 2_{VB}$, detected using the tunnelling spectroscopy as shown in Fig. 1b (filled squares), with the difference between transition II and the bandgap transition I (open squares). The two upper data sets (III) depict the correlation between $\Delta_{CB} = 1P_c - 1S_c$, determined using the tunnelling spectroscopy (filled circles), with the difference between optical transitions III and I (open circles). Also shown is the size dependence of the single-electron charging energy from the tunnelling data (filled triangles).

moving from the transition to the noble metals within a row of the periodic table. In the nanocrystals, as a consequence of the competition between charging and excitation, single-electron tunnelling through the first valence band state exhibits peak multiplicities that vary with size and do not necessarily correlate with the level multiplicity. Nevertheless, Δ_{VB} can be extracted from the wider peak spacing, which is due to both excitation and addition.

The tunnelling data can be further used to decipher the complex QD level structure through a correlation between spacings of levels observed in tunnelling-conductance spectra, with the allowed optical transitions as observed by photoluminescence excitation spectroscopy (PLE) of the InAs nanocrystals⁹. Complexities in the valence band levels of the nanocrystal result from the quantum-confinement-enhanced mixing of close-lying bulk semiconductor bands^{21,22}. The mixing relaxes some of the optical transition selection rules, leading to rich and complex optical spectra of the nanocrystals^{5,9,23}. The above correlation is also important when examining possible effects of charging and tip-induced electric field in the tunnelling measurements on the nanocrystal level structure.

We first compare the size dependence of the bandgap E_g , as extracted from the tunnelling data, with the nanocrystal sizing curve (Fig. 3a). The sizing curve (open diamond data points) was obtained by correlating the average nanocrystal size (that is, radius r), measured using transmission electron microscopy (TEM), with the excitonic bandgap of the same sample^{9,24}. To compare these data with the tunnelling results, we have added a correction term, $1.8e^2/r$, to compensate for the electron–hole excitonic Coulomb interaction which is absent in the tunnelling data⁴ (here e is the electron charge). Agreement is good for the larger nanocrystal radii, with increasing deviation for smaller nanocrystals. The deviation occurs because the TEM sizing curve provides a lower limit to the nanocrystal radius due to its insensitivity to the (possibly amorphous) surface layer. On the other hand, the size extracted from the STM topographic images is overestimated because of the tip–nanocrystal convolution effect²⁰. These differences should be more pronounced for the small sizes, as is indeed observed.

Next, in Fig. 3b, we compare the size dependence of the higher strongly allowed optical transitions with the level spacings measured by tunnelling spectroscopy. We plot excited level spacings versus the observed bandgap for both PLE and tunnelling spectra, thus eliminating the problem of QD size estimation discussed above. The two lower data sets (II) in Fig. 3b compare the difference between the first strong excited optical transition and the bandgap from PLE (E_3-E_1 in ref. 9), with the separation $\Delta_{VB} = 2_{VB} - 1_{VB}$ in the tunnelling data (open and filled squares, respectively). The excellent correlation observed enables us to assign this first excited transition in the PLE to a $2_{VB} - 1S_e$ excitation, as shown schematically in the inset of Fig. 3. Strong optical transitions are allowed only between electron and hole states with the same envelope-function symmetry. We thus infer that the envelope function for state 2_{VB} should have s character and this state can be directly identified as the $2S_{3/2}$ valence band level⁹.

Another important comparison is depicted by the higher pair of curves in Fig. 3b (set III). The second strong excited optical transition relative to the bandgap (E_5-E_1 in ref. 9) is plotted along with the spacing $\Delta_{CB} = 1P_e - S_e$ from the tunnelling spectra. Again, excellent correlation is observed, which allows us to assign this peak in the PLE to the $1_{VB}-1P_e$ transition (Fig. 3, inset). The top-most valence band level, 1_{VB} , should thus have some p character for this transition to be allowed. From this, and considering that the bandgap optical transition $1_{VB}-1S_e$ is also allowed, we conclude that 1_{VB} has mixed s and p character. It is not clear at present whether there is an intrinsic mixing²² or degeneracy between different states having s and p characteristics⁷. Scanning tunnelling spectroscopy experiments in a magnetic field may resolve this issue.

The good correlation between optical transitions and the spacing of levels observed in the tunnelling spectra indicates that the charging did not have a significant effect on the QD level structure. The tunnelling $I-V$ curves are thus well described as an algebraic sum of single electron charging energies and the QD level spacings. □

Received 10 March; accepted 4 June 1999.

1. Alivisatos, A. P. Semiconductor clusters, nanocrystals, and quantum dots. *Science* **271**, 933–937 (1996).
2. Kastner, M. A. Artificial atoms. *Phys. Today* **46**, 24–31 (1993).
3. Ashoori, R. C. Electrons in artificial atoms. *Nature* **379**, 413–419 (1996).
4. Brus, L. E. Electron–electron and electron–hole interaction in small semiconductor crystallites. The size dependence of the lowest excited electronic state. *J. Chem. Phys.* **80**, 4403–4409 (1984).
5. Norris, D. J., Sacra, A., Murray, C. B. & Bawendi, M. G. Measurement of the size dependent hole spectrum in CdSe quantum dots. *Phys. Rev. Lett.* **72**, 2612–2615 (1994).
6. Leon, M., Petroff, P. M., Leonard, D. & Fafard, S. Spatially resolved visible luminescence of self-assembled semiconductor quantum dots. *Science* **267**, 1966–1968 (1995).
7. Gammon, D., Snow, E. S., Shanabrook, B. V., Katzer, D. S. & Park, D. Homogenous linewidths in the optical spectrum of a single gallium arsenide quantum dot. *Science* **273**, 87–90 (1996).
8. Empedocles, S. A., Norris, D. J. & Bawendi, M. G. Photoluminescence spectroscopy of single CdSe nanocrystallite quantum dots. *Phys. Rev. Lett.* **77**, 3873–3876 (1996).
9. Banin, U. *et al.* Size dependent electronic level structure of InAs nanocrystal quantum dots: Test of multiband effective mass theory. *J. Chem. Phys.* **109**, 2306–2309 (1998).
10. Herzberg, G. *Atomic Spectra and Atomic Structure* Ch. 3 (Prentice-Hall, New York, 1937).

11. Guzelian, A. A., Banin, U., Kadavanich, A. V., Peng, X. & Alivisatos, A. P. Colloidal chemical synthesis and characterization of InAs nanocrystal quantum dots. *Appl. Phys. Lett.* **69**, 1432–1434 (1996).
12. Mucic, R. C., Storhoff, J. J., Lestingier, R. L. & Mirkin, C. A. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607–609 (1996).
13. Alivisatos, A. P. *et al.* Organization of nanocrystal molecules using DNA. *Nature* **382**, 609–611 (1996).
14. Colvin, V. L., Schlamp, M. C. & Alivisatos, A. P. Light-emitting diodes made from cadmium selenide nanocrystals and a semiconductor polymer. *Nature* **370**, 354–357 (1994).
15. Klein, D. L., Roth, R., Lim, A. K. L., Alivisatos, A. P. & McEuen, P. L. A single electron transistor made from a cadmium selenide nanocrystal. *Nature* **389**, 699–701 (1997).
16. Colvin, V. L., Goldstein, A. N. & Alivisatos, A. P. Semiconductor nanocrystals covalently bound to metal surface with self-assembled monolayers. *J. Am. Chem. Soc.* **114**, 5221–5230 (1992).
17. Grabert, H. & Devoret, M. H. (eds) *Single Charge Tunneling* (Plenum, New York, 1992).
18. Alpers, B., Cohen, S., Rubinstein, I. & Hodes, G. Room-temperature conductance spectroscopy of CdSe quantum dots using a scanning force microscope. *Phys. Rev. B* **52**, R17017–R17020 (1995).
19. Porath, D., Levi, Y., Tarabia, M. & Millo, O. Tunneling spectroscopy of a single C_{60} molecule in the presence of charging effects. *Phys. Rev. B* **56**, 9829–9833 (1997).
20. Wiesendanger, R. *Scanning Probe Microscopy and Spectroscopy* (Cambridge Univ. Press, 1994).
21. Ekimov, A. I. *et al.* Absorption and intensity-dependent photoluminescence measurements on CdSe quantum dots: assignment of the first electronic transitions. *J. Opt. Soc. Am. B* **10**, 100–107 (1993).
22. Fu, H. & Zunger, A. Excitons in InP quantum dots. *Phys. Rev. B* **57**, R15064–R15067 (1998).
23. Bertram, D., Micic, O. I. & Nozik, A. J. Excited state spectroscopy of InP quantum dots. *Phys. Rev. B* **57**, R4265–R4268 (1998).
24. Banin, U., Lee, J. C., Guzelian, A. A., Kadavanich, A. V. & Alivisatos, A. P. Exchange interaction in InAs nanocrystal quantum dots. *Superlattices Microstruct.* **22**, 559–567 (1997).
25. Alpers, B., Hodes, G., Rubinstein, I., Porath, D. & Millo, O. Energy level tunneling spectroscopy and single electron charging in individual CdSe quantum dots. *Appl. Phys. Lett.* (in the press).

Acknowledgements. We thank B. Alpers, G. Hodes and I. Rubinstein for discussions. This work was supported in part by the Israeli Academy of Sciences and Humanities. U.B. acknowledges the support of an Alon fellowship.

Correspondence and requests for materials should be addressed to O.M. (milode@vms.huji.ac.il).

Surface impact ionization of polar-molecule clusters through pickup of alkali atoms

C. R. Gebhardt, H. Schröder & K.-L. Kompa

Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Strasse 1, 85748 Garching, Germany

The observation that clusters of neutral H_2O (refs 1–4) or SO_2 (ref. 5) molecules, on impact with essentially any solid surface, can decay efficiently into positively and negatively charged fragments has defied explanation, not least because the kinetic energy per molecule can be much smaller than the molecular ionization potentials. Here we present a microscopic model of the charging mechanism, based on a mass analysis of charged SO_2 cluster fragments, which appears to be applicable to polar-molecule clusters more generally. Our mass spectra reveal that all positively charged fragments carry an alkali ion (sodium, potassium or caesium), whereas the negative fragments are simply $(SO_2)_n^-$. The yields of both charged species are comparable, and can be enhanced significantly by pre-treating the sample surface with additional alkali atoms. The key to charge separation in the clusters therefore appears to be the pickup of a neutral (but readily ionized) adatom during impact, followed by delocalization of the adatom's valence electron within the cluster and the subsequent collision-induced fragmentation of the cluster into charged pieces. This process could be of practical use in, for example, charge-pair generation and surface analysis; it may also be relevant to atmospheric ionization processes.

We performed the experiments in a conventional two-chamber molecular-beam apparatus (background pressure, 10^{-7} mbar). Clusters are produced by condensation in a seeded free-jet expansion through a pulsed nozzle (diameter 0.5 mm, typical pulse width 400 μ s, stagnation pressure up to 20 bar). The cluster size distribution is measured with the retarding field technique⁶ employing 30-eV electron impact ionization. The target is mounted in the second chamber perpendicularly to the beam axis, 300 mm downstream from the nozzle. We estimate that a fraction of 2×10^{-4} of the uncondensed particles leaving the nozzle arrive at the surface where

the beam diameter is 8 mm. Most experiments are performed on the SiO_x surface of commercial silicon wafers or on gold (deposited by electrolysis on steel) at a temperature between 400 K and 600 K; that is, under conditions where weakly bound molecular adsorbates have already desorbed. The surface can be pre-treated with alkali-metal atoms by means of an oven source. A plate capacitor arrangement, formed by the target and a grid placed 10 mm in front of it, allows us to measure the total yield of positive and negative charges per pulse. The grid is also used to transfer fragment ions into the detection volume of our pulsed Wiley-McLaren-type time-of-flight mass spectrometer, which has been further optimized for a large extraction volume⁷. The front face of its microchannel plate detector was grounded to ensure that positive and negative ions impinge with the same energy.

Screening experiments have been performed for a variety of clusters striking metallic, as well as insulating, targets. No charge separation is observed for clusters formed by the non-polar molecules O_2 , N_2 , CO_2 , SF_6 , and the noble gases Ne, Xe, Kr. In contrast, clusters of H_2O , SO_2 , NO, NH_3 , NO_2 , SF_4 , CH_3CN , CHClF_2 and isobutene all break into positively and negatively charged fragments on striking the target. This emphasizes that the existence of a permanent molecular dipole moment is required. We chose to examine the SO_2 molecule because it is chemically stable, it lacks complications due to hydrogen bonding and autodissociation⁸ and clusters are readily produced at room temperature. In contrast to the very low electron affinities of small NH_3 or H_2O clusters⁹, the electron affinity of a single SO_2 molecule is ~ 1 eV (ref. 10), which greatly facilitates the formation of stable anion clusters.

By seeding SO_2 in Ne, He and H_2 and applying various reservoir pressures, neutral clusters are produced with variable average sizes $\bar{N}$ and velocities v . From the investigated parameter ranges ($v = 750\text{--}1,750 \text{ m s}^{-1}$, $\bar{N} = 1\text{--}750$) we deduce a threshold velocity

close to 10^3 m s^{-1} ($\bar{N} \approx 300$), and a minimum average cluster size of about $\bar{N} = 20$ ($v \approx 1.2 \times 10^3 \text{ m s}^{-1}$). Comparable thresholds have been reported^{1,2} for clusters of water molecules. The impact of a beam of individual SO_2 molecules ($v \approx 1.2 \times 10^3 \text{ m s}^{-1}$) does not produce any detectable charges.

The four mass spectra of SO_2 cluster fragments, which are partially displayed in Fig. 1, illustrate our explanation of the way in which the impact of a cluster on a surface can lead to the production of positively and negatively charged cluster fragments. The spectra of negatively charged species exhibit a single feature, namely $(\text{SO}_2)_n^-$. The corresponding cations, $(\text{SO}_2)_n^+$ were never observed. Instead, as shown by the assignment, the cations all carry one alkali-metal ion, in this case Na, K or Cs; we have also observed Li. Moreover, pre-treating the surface with, for example, Cs atoms causes a pronounced increase of the corresponding $\text{Cs}^+(\text{SO}_2)_n$ peaks (marked with arrows); it also produces an increase of all the $(\text{SO}_2)_n^-$ peaks, along with an equal increase of the total amount of positive and negative charges. Further treatment of the surface finally leads to a charging efficiency of several per cent per cluster striking the surface. This strong charge-separation signal only gradually disappears after the Cs source has been turned off. This indicates that possible pickup of isolated alkali-metal atoms from the vapour phase is not important, and illustrates the connection between pickup and removal of alkali adatoms from the surface.

We have also measured fragment mass spectra for H_2O , NH_3 and SF_4 . Again, cations carrying an alkali ion are observed, thus confirming the proposed explanation. In the case of NH_3 and SF_4 , we also find indications of intracluster ion-molecule reactions; for H_2O clusters, we find the signature of autoprotolysis. For systems with negligible molecular electron affinity, like water and NH_3 , no (or only weak) negative spectra could be recorded. In such systems,

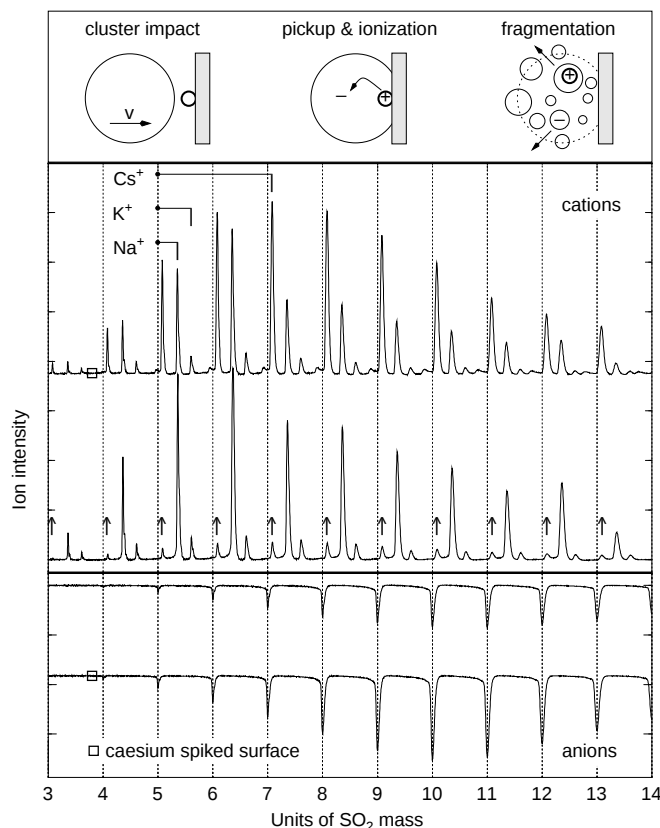


Figure 1 Spectra of positive and negative SO_2 cluster fragments. The clusters have an incident velocity of $1.2 \times 10^3 \text{ m s}^{-1}$, a mean cluster size of 500, and strike a sample consisting of electrolytic gold on steel at 600 K. The mass scale is given in units of SO_2 mass. The cluster cations all carry an additional alkali-metal ion.

These atoms originate from a ubiquitous surface contamination and were picked up from the surface during the course of cluster impact (top). Pre-treating (spiking) the surface with Cs increases the yield of the corresponding $\text{Cs}^+(\text{SO}_2)_n$ progression (marked by arrows) and the overall anion yield.

the shallow Coulomb traps inside our agitated cluster fragments are unable to stabilize the delocalized electron. Hence, the electrons may become detached from the cluster or disappear into the surface. To test this argument, we doped NH_3 clusters with a potential electron scavenger by adding SO_2 as background gas into the nozzle chamber, whereby a pronounced anion spectrum could be recorded. However, reactive intracluster contributions can not be excluded.

Using SO_2 clusters the richest spectrum was observed from a randomly selected pebble, which was used as target without pretreatment. The observation of surface impact ionization in this case shows the ubiquity of alkali-metal atoms on surfaces. Although we are not able to reliably identify the large number of additional peaks, it appears that simultaneous co-pickup of additional neutral species from the surface occurs. Similar rich patterns have been observed from freshly mounted targets held at room temperature. This points to the surface-analytical potential of the surface impact ionization of clusters, which is strictly non-destructive because only adsorbates are picked up.

We also tested the ability of other metal atoms to induce surface impact ionization of clusters by doping the surface via a resistively heated steel loop mounted close to the target. Although *ex situ* surface analysis confirmed the presence of Fe, Ni, Mn and Cr, the observed strong signal was caused exclusively by K and Na atoms. However, an indium-loaded surface seemed indeed to produce the corresponding $(\text{SO}_2)_n^{115}\text{In}^+$ ions, thus indicating that the metal atom needs to have a low ionization potential, rather than the special s^1 electron configuration found in the alkali-metal atoms.

To learn more about the pickup process, the surface was covered with NaCl. Neither strong $\text{Na}^+(\text{SO}_2)_n$ peaks nor any Cl^- carrying anions could be observed. This shows that dissolution of salts does not occur on the timescale of cluster impact, and confirms that atoms in their neutral state are inducing the ionization of fragmenting clusters. In other words, oxidation proceeds much faster than cluster breakup. The general persistence of neutral alkali atoms on surfaces—though surprising—is known from surface analyses applying post-ionization of sputtered neutral particles¹¹.

Metal-atom-doped clusters in the gas phase have frequently been studied^{12–15}. Direct electron delocalization in polar molecular clusters analogous to the liquid phase has been demonstrated for alkali atoms (ref. 16 and references therein; ref. 17), as well as group III metal atoms (ref. 28). Also, accurate molecular dynamics simulations of small solvated structures are available^{18–20}. However, these investigations consider only the equilibrium state reached on a long timescale. Information about the early dynamics of valence-electron delocalization in polar environments has not been available up to now. In the case of surface impact ionization, delocalization has to proceed in the short time window between the cluster–surface contact and the impact-induced fragmentation of the cluster. This window is of the order of the advection time (cluster diameter/cluster velocity), that is, only about one picosecond under our experimental conditions: this is even shorter than (or at most comparable to) electron solvation times^{21,22}. Furthermore, we note that SO_2 (electron affinity $\sim 1\text{ eV}$, dipole moment ~ 1.6 debye) and NH_3 ($\approx 0\text{ eV}$, 1.3 debye) clusters show strong surface impact ionization, although their ability to stabilize an electron is very different²³. From these arguments we conclude that the delocalization of the electron proceeds without a barrier and independently of subsequent ion or electron solvation processes. We propose that the Coulomb fields of the partial point charges associated with the polar environment causes electron delocalization, in a process quite similar to field ionization.

The cluster clearly remains neutral until the collision-induced fragmentation induces the spatial separation of the charge carriers. The energy to overcome the mutual Coulomb attraction is supplied by the kinetic energy of the fragments (inertial charge separation). This concept of examining ion pair formation in neutral clusters through impact fragmentation promises to be useful in the investigation of chemical reactions inside clusters, or cluster-size-depend

ent solvation of salts. In the context of our experiments, the necessity of cluster fragmentation to make charge separation observable rationalizes the existence of a threshold velocity. Also, the existence of a minimum cluster size seems to be quite natural, because a single molecule can neither pick up an atom nor delocalize an electron.

Cluster fragmentation is an issue of general interest^{24,25}. Owing to the varying mean cluster size $\bar{N}$ within the pulse, we found very small fragments ($n \approx 6$) at the beginning but up to $n \approx 100$ towards the end of the supersonic pulse. This indicates a clear trend: large clusters burst into large fragments.

The effect we report here promises to be a new and very efficient addition to the known ionization or electrification mechanisms. We expect that it could be applied in a wide variety of situations; for example, generation of heavy anion plasmas, efficient detection of mesospheric cloud particles, ion propulsion systems, creation of size-selected cluster ions for spectroscopy, or detection and removal of alkali-metal atom contamination from surfaces even in extremely low concentrations.

We note a possible implication of our results for natural electrification processes. The accepted mechanisms for the creation of ions and electrons in the atmosphere are cosmic rays and radioactive decay at the Earth's surface²⁶. Our experimental results suggest that alkali atoms should also be considered as a source of charge carriers, in particular at significant aerosol concentrations. The electrification might occur as follows: a hydrometeor picks up one or several alkali-metal atoms on collision with an aerosol, and increases its intrinsic conductivity through the formation of solvated cations and electrons. This enables one of the many possible cloud-charging mechanisms²⁷ to become effective, or the hydrometeor may directly burst into charged pieces. □

Received 22 March; accepted 22 June 1999.

1. Vostrikov, A. A., Dubov, D. Yu. & Predtechenskiy, M. R. Ionization of water clusters by surface collision. *Chem. Phys. Lett.* **139**, 124–128 (1987).
2. Vostrikov, A. A. & Dubov, D. Yu. Surface induced ionization of neutral water clusters. *Z. Phys. D* **20**, 61–63 (1991).
3. Vostrikov, A. A. *et al.* Ionization of water clusters by collision with surface. *Z. Phys. D* **40**, 542–545 (1997).
4. Andersson, P. U. & Pettersson, J. B. C. Ionization of water clusters by collisions with graphite surfaces. *Z. Phys. D* **41**, 57–62 (1997).
5. Christen, W., Kompa, K.-L., Schröder, H. & Stülpnagel, H. Ionization of SO_2 -clusters by scattering from surfaces. *Ber. Bunsenges. Phys. Chem.* **96**, 1197–1200 (1992).
6. Hagen, O. F. & Obert, W. Cluster formation in expanding supersonic jets: effect of pressure, temperature, nozzle size and test gas. *J. Chem. Phys.* **56**, 1793–1802 (1972).
7. Chandezon, F., Huber, B. & Ristori, C. A new regime Wiley-McLaren time-of-flight mass spectrometer. *Rev. Sci. Instrum.* **65**, 3344–3353 (1994).
8. *Gmelin Handbuch Schwefel*, Ergänzungsband 3, 226 (Springer, Berlin, 1980).
9. Lee, G. H. *et al.* Negative ion photoelectron spectroscopy of solvated electron cluster anions, $(\text{H}_2\text{O})_n^-$ and $(\text{NH}_3)_n^-$. *Z. Phys. D* **20**, 9–12 (1991).
10. Rothe, E. W., Tang, S. Y. & Reck, P. G. Measurement of electron affinities of O_3 , SO_2 , and SO_3 by collisional ionization. *J. Chem. Phys.* **62**, 3829–3831 (1975).
11. Becker, C. H. & Gillen, K. T. Surface analysis of contaminated GaAs: Comparison of new laser-based techniques with SIMS. *J. Vac. Sci. Technol. A* **3**, 1347–1349 (1985).
12. Tang, I. N., Lian, M. S. & Castleman, A. W. Jr Mass spectrometric study of gas-phase clustering reactions: Hydration of the monovalent strontium ion. *J. Chem. Phys.* **65**, 4022–4027 (1976).
13. Donnelly, S. G. & Farrar, J. M. Size-dependent photodissociation cross sections for $\text{Sr}^+(\text{NH}_3)_n$, $n = 3–6$: Rydberg state formation and electron transfer. *J. Chem. Phys.* **98**, 5450–5459 (1993).
14. Weinheimer, C. J. & Lisy, J. M. Vibrational and unimolecular dissociation of mixed solvent cluster ions: $\text{Na}^+((\text{CH}_3)_2\text{CO})_n(\text{CH}_3\text{OH})_m$. *Chem. Phys.* **239**, 357–368 (1998).
15. Ohshima, K., Tsunoyama, H., Yamakita, Y., Misaizu, F. & Ohno, K. Photoionization and density functional study of clusters of alkali metal atoms solvated with acetonitrile molecules, $\text{M}(\text{CH}_3\text{CN})_n$ ($\text{M} = \text{Li}$ and Na). *Chem. Phys. Lett.* **301**, 356–364 (1999).
16. Takasu, R., Misaizu, F., Hashimoto, K. & Fuke, K. Microscopic solvation process of alkali atoms in finite clusters: photoelectron and photoionization studies of $\text{M}(\text{NH}_3)_n$ and $\text{M}(\text{H}_2\text{O})_n$ ($\text{M} = \text{Li}$, Li^+ , Na^+). *J. Phys. Chem. A* **101**, 3078–3087 (1997).
17. Hertel, I. V., Hüglin, C., Nitsch, C. & Schulz, C. P. Photoionization of $\text{Na}(\text{NH}_3)_n$ and $\text{Na}(\text{H}_2\text{O})_n$ clusters: a step towards the liquid phase? *Phys. Rev. Lett.* **67**, 1767–1770 (1991).
18. Barnett, R. N. & Landman, U. Hydration of sodium in water clusters. *Phys. Rev. Lett.* **70**, 1775–1778 (1993).
19. Kim, K. S. *et al.* The nature of a wet electron. *Phys. Rev. Lett.* **76**, 956–959 (1996).
20. Feller, D., Glendenning, E. D., Kendall, R. A. & Peterson, K. A. An extended basis set *ab initio* study of $\text{Li}^+(\text{H}_2\text{O})_n$, $n = 1–6$. *J. Chem. Phys.* **100**, 4981–4997 (1994).
21. Mosyak, A. A., Prezhdo, O. V. & Rossky, P. J. Solvation dynamics of an excess electron in methanol and water. *J. Chem. Phys.* **109**, 6390–6395 (1998).
22. Rips, I. Electron solvation dynamics in polar liquids. *Chem. Phys. Lett.* **245**, 79–84 (1995).
23. Jortner, J. Cluster size effects. *Z. Phys. D* **24**, 247–275 (1992).
24. Christen, W., Even, U., Raz, T. & Levine, R. D. Collisional energy loss in cluster surface impact: Experimental, model, and simulation studies of some relevant factors. *J. Chem. Phys.* **108**, 10262–10273 (1998).
25. Svanberg, M., Ming, L., Markovic, N. & Pettersson, J. B. C. Collision dynamics of large water clusters.

- J. Chem. Phys.* **108**, 5888–5897 (1998).
 26. MacGorman, D. R. & Rust, W. D. *The Electrical Nature of Storms* 32 (Oxford Univ. Press, New York, 1998).
 27. Pruppacher, H. R. & Klett, J. D. *Microphysics of Clouds and Precipitation* 813–814 (Kluwer Academic, Dordrecht, 1996).
 28. Di Palma, T. M., Latini, A., Satta, M. & Giardini Guidoni, A. Molecular beam studies of ammonia clustered with metals produced by pulsed laser reactive ablation. *Int. J. Mass Spectrom.* **179/180**, 319–326 (1998).

Acknowledgements. We thank R. D. Levine for discussions, and H.-J. Schmidtke for contributions to the early stage of this work.

Correspondence and requests for material should be addressed to H.S. (e-mail: hms@mpq.mpg.de).

Triple-isotope composition of atmospheric oxygen as a tracer of biosphere productivity

Boaz Luz*, Eugeni Barkan†, Michael L. Bender†, Mark H. Thieme‡ & Kristie A. Boering§

* The Institute of Earth Sciences, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

† Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA

‡ Department of Chemistry, University of California, San Diego, La Jolla, California 92093, USA

§ Departments of Chemistry and of Geology and Geophysics, University of California, Berkeley, California 94720-1460, USA

Oxygen has three naturally occurring isotopes, of mass numbers 16, 17 and 18. Their ratio in atmospheric O₂ depends primarily on the isotopic composition of photosynthetically produced O₂ from terrestrial and aquatic plants^{1–3}, and on isotopic fractionation due to respiration⁴. These processes fractionate isotopes in a mass-dependent way, such that ¹⁷O enrichment would be approximately half of the ¹⁸O enrichment relative to ¹⁶O. But some photochemical reactions in the stratosphere give rise to a mass-independent isotope fractionation, producing approximately equal ¹⁷O and ¹⁸O enrichments in stratospheric ozone⁵ and carbon dioxide^{6,7}, and consequently driving an atmospheric O₂ isotope anomaly. Here we present an experimentally based estimate of the size of the ¹⁷O/¹⁶O anomaly in tropospheric O₂, and argue that it largely reflects the influences of biospheric cycling and stratospheric photochemical processes. We propose that because the biosphere removes the isotopically anomalous stratosphere-derived O₂ by respiration, and replaces it with isotopically ‘normal’ oxygen by photosynthesis, the magnitude of the tropospheric ¹⁷O anomaly can be used as a tracer of global biosphere production. We use measurements of the triple-isotope composition of O₂ trapped in bubbles in polar ice to estimate global biosphere productivity at various times over the past 82,000 years. In a second application, we use the isotopic signature of oxygen dissolved in aquatic systems to estimate gross primary production on broad time and space scales.

The magnitude of the ¹⁷O anomaly in the present atmosphere can be estimated by comparing $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$ of ambient air O₂ (represented by the HLA standard; see Methods) with O₂ that was not affected by stratospheric processes. To make the latter, we built two airtight terrariums in which O₂ was consumed and replaced biologically. Ultraviolet radiation that could lead to mass-independent fractionation was absent. The terrariums contained *Philodendron* plants, soil and natural water. In both terrariums, O₂ production and consumption occurred via the higher plants as well as via bacteria and algae. Illumination was changed during the experiment with the aim of varying the ratio of photosynthesis to

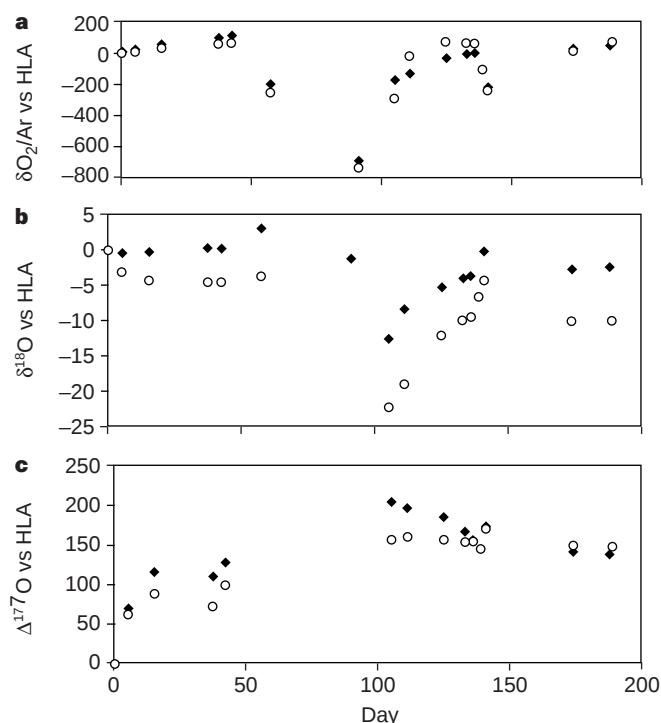


Figure 1 Removal of the atmospheric ¹⁷O anomaly by biological cycling. Shown are variations of $\delta\text{O}_2/\text{Ar}$ (a) $\delta^{18}\text{O}$ (b) and $\Delta^{17}\text{O}$ (c) in the terrarium experiment. Data points: diamonds, terrarium PK; circles, terrarium PDS. Horizontal scale: days from the beginning of the experiment. Light: continuous illumination, days 1–42; room light, days 43–90; 10 h light, 14 h dark, days 91–136 and days 142–198; dark, days 137–141.

respiration, thereby inducing a wide range of $\delta^{18}\text{O}$ and $\delta^{17}\text{O}$ values.

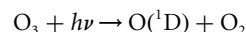
After several turnovers, anomalous ambient air O₂ had been removed by respiration and replaced with normally fractionated O₂, and steady state was attained. Subsequent illumination changes significantly affected O₂ concentration (measured as $\delta\text{O}_2/\text{Ar}$; see Methods) and $\delta^{18}\text{O}$ but not the ¹⁷O anomaly (Fig. 1). The $\delta^{18}\text{O}$ versus $\delta^{17}\text{O}$ trend for data points for which the anomaly was at steady state plot on a nearly perfect straight line ($R^2 = 0.99999$) with a slope of $0.5211 (\pm 0.0005)$, as expected for mass-dependent fractionation⁸ (data not shown). The intercept of the regression line is $0.155 \pm 0.008\text{‰}$ or 155 ± 8 in units of per meg (see Methods). Based on this analysis, we defined the $\Delta^{17}\text{O}$ anomaly as the deviation from normal mass-dependent fractionation ($\Delta^{17}\text{O} = \delta^{17}\text{O} - 0.521\delta^{18}\text{O}$); in the case of the terrarium experiment, $\Delta^{17}\text{O}$ is equal to the intercept value of the regression line. HLA was the preferred standard for high-precision measurements in our study. However, as a reference for $\Delta^{17}\text{O}$ it is admittedly confusing, because it has anomalous isotopic composition. An air sample with no photosynthetic O₂ added will have a $\Delta^{17}\text{O} = 0$ with respect to HLA, and a sample of biologically equilibrated O₂ will have a $\Delta^{17}\text{O} = +155$ per meg with respect to HLA. In this treatment, ocean and meteoric waters are defined as normal with $\Delta^{17}\text{O} = 155$ per meg. A comparison of the isotopic composition of ocean water (represented by V-SMOW) to air O₂ ($\delta^{18}\text{O} = -22.960\text{‰}$ and $\delta^{17}\text{O} = -11.778\text{‰}$)⁶, supports our conclusions that air bears a mass-independent signature. The $\Delta^{17}\text{O}$ of V-SMOW is calculated as 184 per meg [$\Delta^{17}\text{O} = (-11.778 - 0.521 \times -22.96) \times 1,000$]. In this calculation, the derived anomaly is very sensitive to the slope term (0.521), and thus the small difference between 184 and 155 per meg may not be significant.

Importantly, the terrarium experiment cannot represent all Earth-surface processes affecting $\Delta^{17}\text{O}$. For example, the range of isotopic variations in global meteoric waters is much greater than in

the experiment, and a recent study⁹ indicates that in these waters the $\delta^{17}\text{O}/\delta^{18}\text{O}$ slope is slightly different than in respiration. Furthermore, the humidity in the terrariums was at saturation and important isotope fractionation due to evapo-transpiration from leaves^{2,3} was not reflected in our experiment. In addition, the $\delta^{17}\text{O}/\delta^{18}\text{O}$ slopes in the various processes consuming oxygen (dark respiration, cyanide resistant respiration, photorespiration and the Mehler reaction) may vary slightly from 0.521. Because the relative rates of these processes in our experiment are not expected to be the same as in the global biosphere, the intercept of the regression line in natural systems may differ from the value we measured. Thus, while the 155 per meg estimate clearly demonstrates the anomalous isotopic signature of atmospheric O_2 , further study is needed in order to better constrain the magnitude of the anomaly.

Our experimental determination of the mass-independent anomaly in O_2 can be compared with estimates of its stratospheric production. Bender *et al.*¹⁰ first suggested that photochemical mass-independent fractionation processes in the stratosphere involving

ozone and CO_2 species could result in anomalous O_2 . In the stratosphere, the ozone recombination reaction, $\text{O} + \text{O}_2 \rightarrow \text{O}_3$, causes O_3 to be mass-independently fractionated. The anomalous fractionation is well documented, although its cause is debated^{11–13}. Theory¹⁴ and laboratory experiments¹⁵ suggest that the anomalous ozone enrichment is transferred to CO_2 . Ultraviolet photolysis of ozone in the stratosphere generates an electronically excited oxygen atom which can undergo isotope exchange with CO_2 :



Thus, because the ultimate source of the oxygen in O_3 and $\text{O}(^1\text{D})$ in the stratosphere is the O_2 reservoir, O_2 becomes anomalously depleted as CO_2 becomes anomalously enriched. We note that there is no stratospheric loss term for the enrichment in CO_2 , and the net stratospheric enrichments are lost only at the Earth's surface by isotope exchange with liquid water in leaves and in the ocean^{3,14,16} (Fig. 2). In contrast, O_2 does not exchange isotopes with water¹⁷, and the depletion disappears only through the consumption of O_2 by respiration and its replacement by photosynthesis. The respiratory and photosynthetic fluxes are relatively small (Fig. 2) compared with the stratospheric production, and thus the $\Delta^{17}\text{O}$ anomaly accumulates to a measurable level of 155 per meg over the residence time of atmospheric O_2 ($\sim 1,200$ yr; ref. 10).

To test the hypothesis that stratospheric processes can generate the 155 per meg anomaly in atmospheric O_2 , we estimate the production rate of anomalous O_2 from stratospheric photochemistry. We use observations of mass-independently fractionated CO_2 and its correlation with N_2O (a long-lived tracer that is photolysed in the stratosphere), coupled with calculations of the annual mass flux of air from the stratosphere to the troposphere. We frame the calculations in terms of N_2O for two reasons. First, observations of $\Delta^{17}\text{O}$ of CO_2 ($\Delta^{17}\text{O}_{\text{CO}_2}$) are extremely sparse, so their tight correlation with N_2O —a species that has been extensively measured in the stratosphere—serves as a proxy for the distribution of $\Delta^{17}\text{O}_{\text{CO}_2}$ in the stratosphere. Second, we can use knowledge of either the stratospheric loss rate for N_2O (ref. 18) or the age of stratospheric air as a function of N_2O (ref. 19) in order to calculate the N_2O mixing ratio, and therefore $\Delta^{17}\text{O}_{\text{CO}_2}$, in air returning to the troposphere from the stratosphere. Table 1 shows annual production rates of $\Delta^{17}\text{O}_{\text{CO}_2}$ estimated from two different mass-flux calculations^{20,21} coupled with the two means of estimating N_2O returning to the troposphere. In order to estimate the anomalous O_2 transferred to the troposphere each year, we assume that stratospheric CO_2 is the only species that can sequester a mass-independent enrichment that leaves the O_2 reservoir anomalously depleted. We also assume that photochemical transfer of the mass-independent fractionation from O_3 to CO_2 via $\text{O}(^1\text{D})$ scales with both CO_2 and O_3 abundances. Because the O_2 anomaly accumulates over 1,200 years, knowledge of CO_2 and O_3 levels back to pre-industrial times are required. While CO_2 levels are well constrained from ice-core data, we must rely on photochemical–dynamical atmospheric models for pre-industrial ozone levels. Crutzen and Bruhl²² and Martiniere *et al.*²³ predict pre-industrial stratospheric ozone levels greater than the present-day atmosphere by up to 30–40% over significant regions of the stratosphere. The annual production rate (P) of anomalous O_2 ($\Delta^{17}\text{O}_{\text{O}_2}$) is then calculated from:

$$P[\Delta^{17}\text{O}_{\text{O}_2}](\text{per meg yr}^{-1}) = -fP[\Delta^{17}\text{O}_{\text{CO}_2}](\text{per meg yr}^{-1}) \quad (1)$$

The factor f allows scaling for the pre-industrial ratio of CO_2/O_2 (280/210,000), for the increase of CO_2 to its present level (358 p.p.m.) and for a $\sim 30\%$ decrease in O_3 since pre-industrial times. Thus, $f = (280/210,000)(280/358)1.3$. The estimated 1,200–

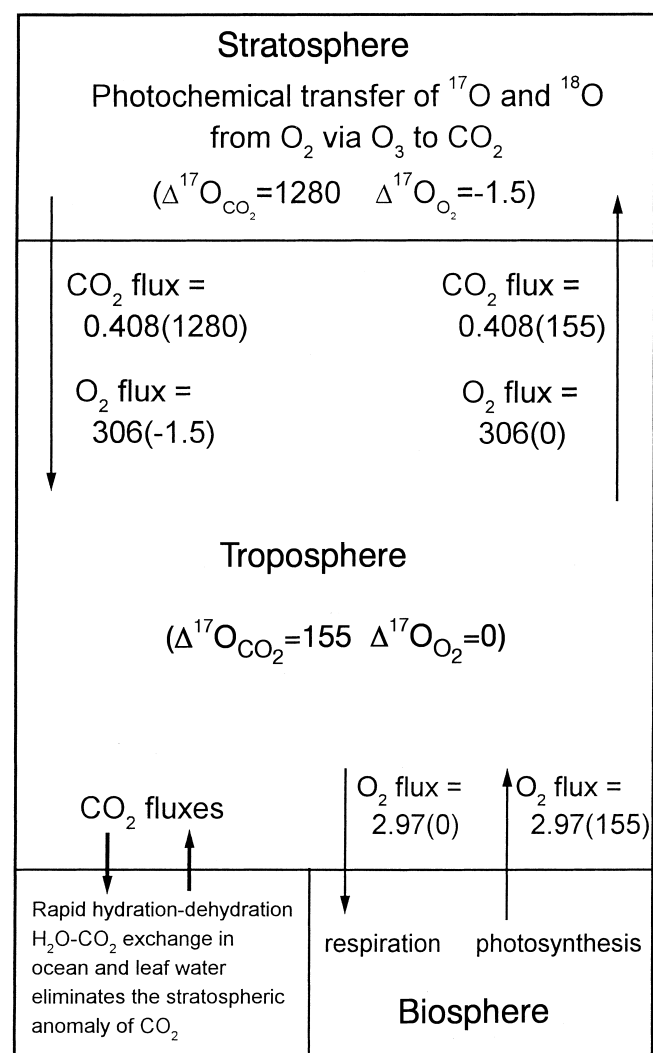


Figure 2 Simplified O_2 , CO_2 and $\Delta^{17}\text{O}$ cycle. The values of $\Delta^{17}\text{O}_{\text{O}_2}$ and $\Delta^{17}\text{O}_{\text{CO}_2}$ (in units of per meg with respect to HLA) are indicated for the stratosphere and the troposphere. The fluxes shown are in units of $10^{16} \text{ mol yr}^{-1}$ and the values in parenthesis are $\Delta^{17}\text{O}_{\text{O}_2}$ and $\Delta^{17}\text{O}_{\text{CO}_2}$. The tropospheric value of $\Delta^{17}\text{O}_{\text{CO}_2}$ is taken as equal to $\Delta^{17}\text{O}$ of ocean and leaf water (155 per meg, see text). Note that the net positive anomaly flux in CO_2 ($1280 \times 0.408 - 155 \times 0.408$) from stratosphere to the troposphere is of the same absolute magnitude as the negative anomaly flux in O_2 (-1.50×306). This negative anomaly flux from the stratosphere is balanced by the positive anomaly flux in O_2 from the biosphere (155×2.97).

Table 1 Calculation of the $\Delta^{17}\text{O}$ of atmospheric O_2

Mass flux of air from the stratosphere ($10^{17} \text{ kg yr}^{-1}$)	Stratospheric turnover time (yr)	$[\text{N}_2\text{O}]$ returning to troposphere (p.p.b.v.)	$\Delta^{17}\text{O}_{\text{CO}_2}$ (per meg) [#]	$P[\Delta^{17}\text{O}_{\text{CO}_2}]$ (per meg yr^{-1})	$P[\Delta^{17}\text{O}_{\text{O}_2}]$ (per meg yr^{-1})	Estimated 1,200-yr accumulation of $\Delta^{17}\text{O}_{\text{O}_2}$ (per meg) ^{††}
2.0*	—	246	1,550	67 [‡]	−0.091	−109
6.8†	—	292	530	80 [‡]	−0.109	−131
—	2.6‡	250 [¶]	1,460	63 ^{**}	−0.086	−104
—	1.0§	294 [¶]	500	76 ^{**}	−0.103	−124

* Mass flux across the 100-mbar surface (based on ref. 20).

† Mass flux across the 380 K potential-temperature surface (based on ref. 21).

‡ Stratosphere defined as mass of atmosphere above 100 mbar ($5.2 \times 10^{17} \text{ kg}$, based on ref. 20).

§ Stratosphere defined as mass of atmosphere above 380 K ($6.8 \times 10^{17} \text{ kg}$, based on ref. 21).

|| Calculated using an N_2O loss rate of $12.5 \text{ Mt N yr}^{-1}$ and a growth rate of 0.3 Mt N yr^{-1} (ref. 18), and mass fluxes of air into and out of the stratosphere in column 1: flux out = flux in − loss rate − growth rate.

¶ $[\text{N}_2\text{O}]$ for stratospheric air with a mean age equivalent to the turnover time in column 2, derived from simultaneous observations of stratospheric CO_2 and N_2O mixing ratios¹⁸.

Extrapolating from observations^{6,7,28} of the correlation of $\Delta^{17}\text{O}_{\text{CO}_2}$ with N_2O mixing ratios (or CH_4 and the relation between stratospheric CH_4 and N_2O) to the values of N_2O in column 3.

** Annual production rate of $\Delta^{17}\text{O}_{\text{CO}_2}$ in the troposphere by transport from the stratosphere, including dilution factors (mass flux from stratosphere/mass of the troposphere) of 0.043 and 0.153, respectively, for the mass fluxes in column 1.

†† Note that using a more realistic time-varying CO_2 time series over the past 1,200 years, and integrating CO_2 rather than using a constant f in equation (1), alters these estimates by <2%.

Note also that a value of 1.3 was used (included in f) to scale $[\text{O}_3]$ back to pre-industrial times, a number that may represent an upper limit based on model predictions in ref. 23 for which changes in this magnitude were predicted above 35 km and below 25 km.

year accumulation of anomalous O_2 from stratospheric photochemical fractionation from these production rates ranges from 104 to 131 per meg (Table 1), consistent with the measured anomaly of 155 per meg. Carrying the known uncertainties through these calculations yields an error estimate of $\pm 25\%$, and the true uncertainty may be higher given our present state of knowledge of mass-independent fractionation processes in the stratosphere and of rates of stratosphere–troposphere exchange over the past 1,200 years.

Additional confirmation of the stratospheric origin of the atmospheric anomaly comes from isotope mass balance in the troposphere (Fig. 2). Anomalous CO_2 with excess ^{17}O enters the troposphere and brings a net anomalous flux $([1,280 \times 0.408 - 155 \times 0.408]10^{16} \text{ per meg mol yr}^{-1})$. A parallel negative anomaly flux of the same absolute magnitude is carried by O_2 ($-1.5 \times 306 \times 10^{16} \text{ per meg mol yr}^{-1}$). The anomaly in the CO_2 flux is removed by hydration–dehydration exchange with water, and the negative flux in O_2 is balanced by a positive anomaly flux from the biosphere ($155 \times 2.97 \times 10^{16} \text{ per meg mol yr}^{-1}$). The value of $\Delta^{17}\text{O}_{\text{CO}_2}$ in air returning from the stratosphere was calculated as $1,280 \text{ per meg}$ ($-\Delta^{17}\text{O}_{\text{O}_2}/f + 155$, with f as in equation (1)). This estimate is in good agreement with the corresponding values in Table 1. Thus, the two independent mass-balance calculations strongly suggest that the magnitude of the atmospheric $\Delta^{17}\text{O}_{\text{O}_2}$ anomaly reflects the ratio between two important global processes—biospheric O_2 production and stratospheric photochemistry involving O_2 , O_3 and CO_2 .

In an attempt to learn about past variations in global biosphere production, we have analysed ice-core samples from Summit, Greenland (GISP2 ice core, Table 2). Here we make a preliminary and provisional interpretation of the results using the following equation:

$$\Delta^{17}\text{O}_{\text{O}_2}^* = k[\text{CO}_2]/P \quad (2)$$

where $\Delta^{17}\text{O}_{\text{O}_2}^* = \Delta^{17}\text{O}_{\text{O}_2} - 155 \text{ per meg}$, $[\text{CO}_2]$ is the atmospheric CO_2 concentration, k is a proportionality constant relating the rate of anomalous O_2 production to $[\text{CO}_2]$, and P is gross production by the global biosphere. In writing this equation, we assume that the production rate of anomalously depleted O_2 in the stratosphere is proportional to the CO_2 mixing ratio, and that k is constant between glacial and interglacial times. Of course, k enfold contributions from both the stratospheric circulation and isotope photochemistry involving ultraviolet flux, O_3 and CO_2 , and there are significant uncertainties in predicting these aspects of past atmospheres. Crutzen and Bruhl²² and Martinieri *et al.*²³ predict that perturbations to stratospheric ozone from changes in temperature and chemistry (for example, due to changes in $[\text{N}_2\text{O}]$ and $[\text{CH}_4]$) between the pre-industrial interglacial and glacial times largely cancel, with modelled ozone changes at all stratospheric altitudes of 10% or less. Changes in the stratospheric circulation due to dynamical feedbacks as climate has changed appear to be more uncertain²⁴. However, it is not unreasonable to expect that the effect of changes in the circulation on the global k could largely cancel out. If the circulation slowed down, the $\Delta^{17}\text{O}$ anomalies of air returning from the stratosphere would be larger but with a smaller mass flux, whereas if the circulation strengthened, the anomalies would be smaller but with a larger mass flux into the troposphere. More complicated chemical/dynamical/biological ozone feedbacks might also have occurred; these include changes in tropopause height which might alter stratospheric water vapour, or changes in the ocean biospheric production of OCS which is oxidized to sulphuric acid in the stratosphere, producing aerosol. Until further modelling studies of stratospheric chemistry and dynamics in past atmospheres are made, we assume that k is constant. We then calculate 'normalized gross biosphere production', P_t/P_o , as follows, where the subscript t denotes time before present and subscript o denotes present:

$$P_t/P_o = k_t/k_o([\text{CO}_2]_t/[\text{CO}_2]_o) \times (\Delta^{17}\text{O}_{\text{O}_2}^*)_o/(\Delta^{17}\text{O}_{\text{O}_2}^*)_t \quad (3)$$

Table 2 Isotope data and normalized productivity in the GISP2 ice core

Depth (m)	Gas age (kyr)	Climate	$\delta^{15}\text{N}$ (‰)	$\delta^{18}\text{O}$ (‰)*	$\delta^{17}\text{O}$ (‰)*	$\Delta^{17}\text{O}_{\text{O}_2}$ (per meg)	$[\text{CO}_2]$ (p.p.m.v.)†	P_t/P_o ‡
1,907	0.15	Interglacial	0.41	1.050	0.584	0	280	1.00
2,039	18.67	Glacial	0.41	1.050	0.584	37	190	0.89
2,039	25.78	Glacial	0.36	0.822	0.467	39	190	0.91
2,212	36.67	Interstadial	0.39	0.413	0.236	21	210	0.87
2,492	56.17	Interstadial	0.42	0.365	0.226	35	210	0.97
2,668	82.03	Interglacial	0.32	0.236	0.135	12	235	0.91

* The oxygen isotope data shown were corrected for gravitational fractionation²⁷ by subtraction of the $\delta^{15}\text{N}$ value from the measured $\delta^{17}\text{O}$ and by subtraction of $2(\delta^{15}\text{N})$ from the measured $\delta^{18}\text{O}$.

† From ref. 29.

‡ Normalized productivity.

Normalized gross-production values, calculated for 5 times during the past 82 kyr, range from 0.87 to 0.97, not greatly different from the present (Table 2). Since terrestrial-biosphere production during glacial times was probably somewhat lower than today (due to glacier-ice advance, colder temperatures, lower p_{CO_2} and drier climates), ocean productivity was apparently similar to or higher than today. For example, samples from the Last Glacial Maximum have normalized gross production of 0.9. Meyer²⁵ estimated that net productivity of the land-biosphere was 0.75 times today's. Terrestrial gross production of O_2 may have been a few per cent higher, since photorespiration would have been proportionally faster in response to lower CO_2 (D. Yakir, personal communication). Oceanic gross-production, which today accounts for ~40% of global photosynthesis, would then have been slightly higher than today's value.

In a similar way to the application of $\Delta^{17}\text{O}$ as a tracer of global biospheric production, $\Delta^{17}\text{O}$ of dissolved O_2 ($\Delta^{17}\text{O}_{\text{diss.O}_2}$) can be used for inferring the rate of O_2 production by aquatic organisms. The $\Delta^{17}\text{O}_{\text{diss.O}_2}$ in the photic zone of oceans and lakes depends on the relative rates of air–water gas exchange (which introduces the stratospheric anomaly to dissolved O_2) and biological- O_2 cycling (which removes this anomaly). Our measurements of $\Delta^{17}\text{O}_{\text{diss.O}_2}$ over a range of biological productivity demonstrate this point. In the Dead Sea, where at present photosynthesis is zero, $\Delta^{17}\text{O}_{\text{diss.O}_2}$ with respect to HLA is –7 per meg. This value, as expected, is identical within the analytical error to atmospheric oxygen. In the Red Sea (Station A near Eilat, 10 m depth, May 1998), $\Delta^{17}\text{O}_{\text{diss.O}_2}$ was 51 per meg, indicating moderate production. In the highly productive Sea of Galilee (3 m depth; February–April 1998), $\Delta^{17}\text{O}_{\text{diss.O}_2}$ values ranged from 114 to 136 per meg and clearly demonstrate the effect of rapid biological production. Infinitely rapid production would completely eliminate the anomaly, and $\Delta^{17}\text{O}_{\text{diss.O}_2}$ would equal $\Delta^{17}\text{O}$ of H_2O (taken as 155 per meg).

At the time of our sampling in the Sea of Galilee, the rate of air–lake gas exchange was about $0.15 \text{ mol m}^{-2} \text{ d}^{-1}$ (based on the dependence of gas exchange in lakes on wind speed²⁶). In April 1998, $\Delta^{17}\text{O}_{\text{diss.O}_2}$ was 136 per meg and the lake thus lost 136×0.15 per meg mol d^{-1} to the atmosphere. Assuming for simplicity that the lake was at steady state, this loss was balanced by gross production (P) anomaly flux of 155P per meg mol d^{-1} , and P can be calculated as 0.13 mol d^{-1} ($136/155 \times 0.15$). This example shows that if the rate of air–water gas exchange is known, $\Delta^{17}\text{O}_{\text{diss.O}_2}$ can be used to constrain gross production. Constraining production in this way has the advantage that it reflects a spatially and temporally integrated value, and can be easily applied over wide regions in lakes and seas. □

Methods

In the terrarium experiment, water was introduced to the bottom of the chambers through tubing connected to an outside reservoir made of flexible Tedlar in order to compensate for gas gains and losses and to keep constant ambient pressure. The light sources were fluorescent lamps (about $100 \mu\text{E m}^{-2} \text{ s}^{-1}$ with negligible ultraviolet flux). Terrarium PK contained water from the Sea of Galilee ($\delta^{18}\text{O} = -0.5\text{‰}$ vs V-SMOW), and terrarium PDS contained Dan River water ($\delta^{18}\text{O} = -6.6\text{‰}$ vs V-SMOW). About 1.5 ml of dissolved gases were extracted from water samples, and about 3 ml air were extracted from ice-core samples. Most mass spectrometric analyses were carried out at the Hebrew University of Jerusalem. Dried and CO_2 -free air was passed with a high purity He carrier (30 ml min^{-1}) through a chromatographic column ($5 \text{ m} \times 2 \text{ mm}$ inside diameter stainless-steel tube packed with 45/60 mesh 5-Å molecular sieve and held at 0°C). O_2 and Ar eluted completely in 15 min, and were collected by passing the carrier gas through a trap containing coarse 5-Å molecular sieve held at -196°C . The remaining He carrier was pumped out, the purified O_2 and Ar were transferred to stainless-steel tubes²⁷ and admitted to a multi-collector mass spectrometer (Finnigan Delta-Plus). The $\delta^{18}\text{O}$ and $\delta^{17}\text{O}$ of O_2 and $\delta\text{O}_2/\text{Ar}$ were measured against a reference O_2/Ar mixture that was

calibrated against air standard HLA. This standard was prepared by cryogenic drying (-80°C) of 4 litres of outside air and is stored in a stainless-steel container. Because atmospheric mixing is rapid compared to the processes altering its isotopic composition, HLA is representative of all atmospheric oxygen. In all cases, we analysed samples in duplicate. Corrections were applied in order to account for the sensitivity of ionization efficiencies of the three isotope species of oxygen to variations in the O_2/Ar ratio. Isotopic and O_2/Ar ratios are given as: $\delta\text{O}_2/\text{Ar}(\text{‰}) = [(\text{O}_2/\text{Ar}_{\text{sample}})/(\text{O}_2/\text{Ar}_{\text{HLA}}) - 1]10^3$; $\delta^*\text{O}(\text{‰}) = (^*\text{O}/^{16}\text{O}_{\text{sample}})/(^*\text{O}/^{16}\text{O}_{\text{HLA}}) - 1]10^3$ where $^*\text{O}$ denotes ^{17}O or ^{18}O ; and $\Delta^{17}\text{O}$ (per meg) = $(\delta^{17}\text{O} - 0.521\delta^{18}\text{O})10^3$. The analytical precision (standard error) of $\delta^{18}\text{O}$, $\delta^{17}\text{O}$ and $\Delta^{17}\text{O}$ measurements were 0.003‰, 0.009‰ and 9 per meg, respectively. For the purpose of cross calibration, standards and duplicates of ice-core samples were run also at UCSD. The average inter-laboratory differences for $\delta^{18}\text{O}$, $\delta^{17}\text{O}$ and $\Delta^{17}\text{O}$ were 0.044‰, 0.029‰ and 7 per meg, respectively.

Received 20 November 1998; accepted 8 June 1999.

- Guy, R. D., Fogel, M. L. & Berry, J. A. Photosynthetic fractionation of stable isotopes. *Plant. Physiol.* **101**, 37–47 (1993).
- Dongmann, G. The contribution of land photosynthesis to the stationary enrichment of the ^{18}O in the atmosphere. *Radiat. Environ. Biophys.* **11**, 219–255 (1974).
- Farquhar, G. D. et al. Vegetation effects on the isotope composition of oxygen in atmospheric CO_2 . *Nature* **363**, 439–443 (1993).
- Lane, G. & Dole, M. Fractionation of oxygen isotopes during respiration. *Science* **123**, 574–576 (1956).
- Schueler, B., Morton, J. & Mauersberger, K. Measurement of isotopic abundances in collected stratospheric ozone samples. *Geophys. Res. Lett.* **17**, 1295–1298 (1990).
- Thiemens, M. H., Jackson, T., Zipf, E. C., Erdman, P. W. & van Egmond, C. Carbon dioxide and oxygen isotope anomalies in the mesosphere and stratosphere. *Science* **270**, 969–972 (1995).
- Thiemens, M. H., Jackson, T. L. & Brenninkmeijer, C. A. M. Observation of a mass independent oxygen isotopic composition in terrestrial stratospheric CO_2 , the link to ozone chemistry, and the possible occurrence in the Martian atmosphere. *Geophys. Res. Lett.* **22**, 225–257 (1995).
- Thiemens, M. H. Mass-independent isotope effects in planetary atmospheres and the early solar system. *Science* **283**, 341–345 (1999).
- Li, W. J. & Meijer, H. A. J. The use of electrolysis for accurate $\delta^{17}\text{O}$ and $\delta^{18}\text{O}$ isotope measurements in water. *Isotopes Environ. Health Studies* **34**, 349–369 (1998).
- Bender, M., Sowers, T. & Labeyrie, L. The Dole effect and its variations during the last 130,000 years as measured in the Vostok ice core. *Global Biogeochem. Cycles* **8**, 363–376 (1994).
- Thiemens, M. H., Jackson, T., Mauersberger, K., Schueler, B. & Morton, J. Oxygen isotope fractionation in stratospheric CO_2 . *Geophys. Res. Lett.* **18**, 669–672 (1991).
- Thiemens, M. H. & Jackson, T. Pressure dependency for heavy isotope enhancement in ozone formation. *Geophys. Res. Lett.* **17**, 717–719 (1990).
- Mauersberger, K., Erbacher, B., Kranowsky, D., Gunther, J. & Nickel, R. Ozone isotope enrichment: Isotopomer-specific rate coefficients. *Science* **283**, 370–372 (1999).
- Yung, Y. L., Lee, A. Y. T., Irion, F. W., DeMore, W. B. & Wen, J. Carbon dioxide in the atmosphere: Isotopic exchange with ozone and its use as a tracer in the middle atmosphere. *J. Geophys. Res.* **102**, 10857–10866 (1997).
- Wen, J. & Thiemens, M. H. First multi-isotope study of the $\text{O}(^1\text{D}) + \text{CO}_2$ exchange and stratospheric consequences. *J. Geophys. Res.* **98**, 12801–12808 (1993).
- Francey, R. J. & Tans, P. P. Latitudinal variations in oxygen-18 of atmospheric CO_2 . *Nature* **327**, 495–497 (1987).
- Dole, M. & Jenks, G. Isotopic composition of photosynthetic oxygen. *Science* **100**, 409 (1944).
- Minschwaner, K., Salawitch, R. J. & McElroy, M. B. Absorption of solar radiation by O_3 : Implications for O_3 and lifetimes of N_2O , CFCl_3 , and CF_2Cl_2 . *J. Geophys. Res.* **98**, 10543–10561 (1993).
- Boering, K. A. et al. Stratospheric mean ages and transport rates from observations of carbon dioxide and nitrous oxide. *Science* **274**, 1340–1343 (1996).
- Holton, J. R. On the global exchange of mass between the stratosphere and the troposphere. *J. Atmos. Sci.* **47**, 392–395 (1990).
- Appenzeller, C., Holton, J. R. & Rosenlof, K. H. Seasonal variation of mass transport across the tropopause. *J. Geophys. Res.* **101**, 15071–15078 (1996).
- Crutzen, P. J. & Bruhl, C. A model study of atmospheric temperatures and the concentrations of ozone, hydroxyl, and some other photochemically active gases during the glacial, the pre-industrial Holocene and the Present. *Geophys. Res. Lett.* **20**, 1047–1050 (1993).
- Martinerie, P., Brasseur, G. P. & Granier, C. The chemical composition of ancient atmospheres: A model study constrained by ice core data. *J. Geophys. Res.* **100**, 14291–14304 (1995).
- Rind, D. & Lacis, A. The role of the stratosphere in climate change. *Surv. Geophys.* **14**, 133–165 (1993).
- Meyer, M. K. *Net Primary Productivity Estimates for the Last 18,000 years Evaluated from Simulations by a Global Climate Model*. Thesis, Univ. Wisconsin (1988).
- Clark, J. F. et al. in *Air–Water Gas Transfer* (eds Jaehne, B. & Monahan, E. C.) 785–800 (Aeon, Hanau, 1995).
- Sowers, T., Bender, M. & Raynaud, D. Elemental and isotopic composition of occluded O_2 and N_2 in polar ice. *J. Geophys. Res.* **94**, 5137–5150 (1989).
- Brenninkmeijer, C. A. M., Lowe, D. C., Manning, M. R., Sparks, R. J. & van Velthoven, P. F. J. The ^{13}C , ^{14}C and ^{18}O isotopic composition of CO , CH_4 , and CO_2 in the higher southern latitudes lower stratosphere. *J. Geophys. Res.* **100**, 26163–26172 (1995).
- Barnola, J. M., Pimienta, P., Raynaud, D. & Korotkevich, Y. S. CO_2 –climate relationship as deduced from the Vostok ice core—A reexamination based on new measurements and on a reevaluation of the air dating. *Tellus B* **43**, 83–90 (1991).

Acknowledgements. We appreciate the help of Y. Yacobi, and thank J. Orchard and Y. Sagi for help with sample preparation. We thank the USA–Israel BSF, The Israel Science Foundation and the Moshe-Shilo Minerva Center for support; we also thank the Office of Polar Programs of the NSF, and the National Institute of Global Environmental Change, Department of Energy for their support of the ice-core study. M.H.T. thanks the NSF for support.

Correspondence and requests for materials should be addressed to B.L. (e-mail: boazluz@cc.huji.ac.il).

Relative influences of atmospheric chemistry and transport on Arctic ozone trends

M. P. Chipperfield* & R. L. Jones†

* The Environment Centre, University of Leeds, Leeds LS2 9JT, UK

† Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

The reduction in the amount of ozone in the atmospheric column over the Arctic region, observed during the 1990s^{1,2}, resembles the onset of the Antarctic ozone 'hole' in the mid-1980s, but the two polar regions differ significantly with respect to the relative contributions of chemistry and atmospheric dynamics to the ozone abundance. In the strong, cold Antarctic vortex, rapid springtime chemical ozone loss occurs throughout a large region of the lower stratosphere, whereas in the Arctic, although chemical ozone depletion has been observed^{3–11}, the vortex is generally much smaller, weaker and more variable¹². Here we report a model-based analysis of the relative importance of dynamics and chemistry in causing the Arctic ozone trend in the 1990s, using a state-of-the-art three-dimensional stratospheric chemistry–transport model. North of 63°N we find that, on average, dynamical variations dominate the interannual variability, with little evidence for a trend towards more wintertime chemical depletion. However, increases in the burden of atmospheric halogens since the early 1970s are responsible for a large (14%) reduction in the average March column ozone, but this effect is mostly caused by increased destruction throughout the year rather than by halogen chemistry associated with wintertime polar stratospheric clouds. Any influence of climate change on future average Arctic ozone amounts may thus be dominated by possible circulation changes, rather than by changes in chemical loss.

Early observations of stratospheric ozone indicated the important role of transport, as well as chemistry, in determining the distribution of stratospheric ozone¹³. Recent studies have indicated that changes in atmospheric circulation may have contributed to observed changes in mid-latitude ozone from 1979 to 1991^{14,15}. Three-dimensional (3D) chemistry–transport models (CTMs) are now able to represent the seasonal chemistry and transport of stratospheric trace species. These models^{16–18} specify meteorological analyses for the winds and temperatures, allowing the calculated distribution of atmospheric trace species to be compared directly with measurements. We have now updated our 3D CTM to be suitable for multiannual integrations¹⁹, an advance which allows us to model the interannual variability in stratospheric dynamics and chemistry in a single simulation. We integrated the CTM using a horizontal resolution of 7.5° × 7.5°, and 12 potential-temperature levels from 335 K to 2,700 K (resolution of ~3 km in the lower stratosphere). The start of the simulation was October 1991 using horizontal winds and temperatures from the UK Meteorological Office²⁰. The CTM calculates the vertical (adiabatic) motion from a radiation scheme. The standard CTM includes a description of heterogeneous chemistry occurring on liquid/solid polar stratospheric clouds (PSCs) and mid-latitude sulphate aerosols¹⁹. The standard CTM simulations used a constant halogen loading of 3.6 parts per billion by volume (p.p.b.v.) of total chlorine and 20 parts per trillion (10¹²) by volume (p.p.t.v.) of total bromine.

Newman *et al.*¹ reported satellite data which show that the March monthly-mean ozone-column values in the region 63–90°N

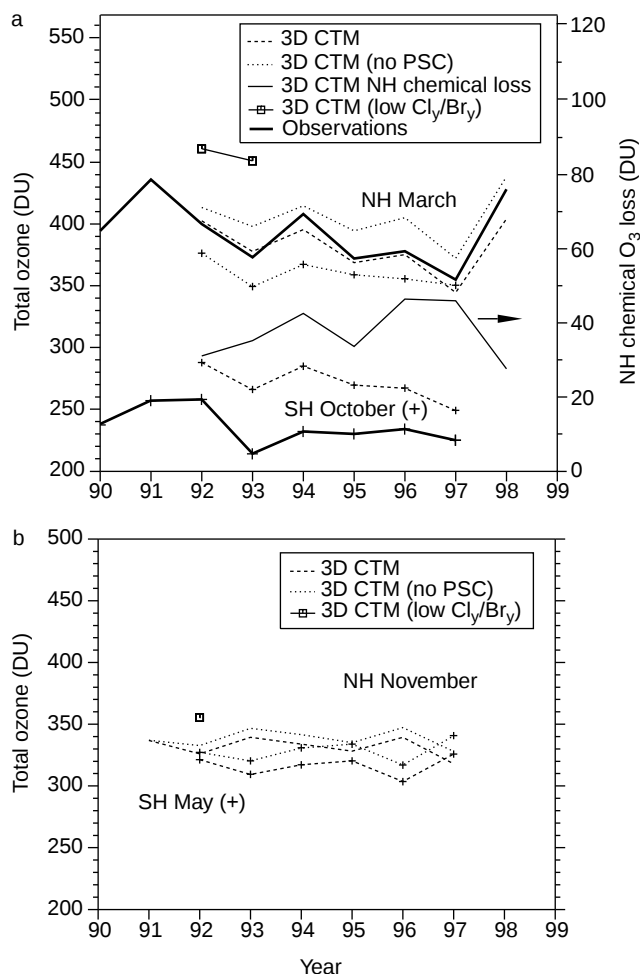


Figure 1 Monthly average column ozone (DU) between 63° and 90° latitude. **a**, March in the Northern Hemisphere (NH) and October in the Southern Hemisphere (SH); **b**, November in the NH and May in the SH. Panel **a** also shows the satellite observations of Newman *et al.*¹ extended until March 1999. Results are shown from three 3D model simulations. The standard model (dashed line) uses present-day halogen loadings and includes the effects of heterogeneous chemistry on polar stratospheric clouds (PSCs). The model includes a time-dependent distribution of liquid aerosols which is taken from 2D model calculations. The dotted line shows results of the simulation without the effects of chlorine activation on liquid and solid PSCs and aerosols. The squares show the NH results of the model simulation with lower halogen loadings (1.35 p.p.b.v. chlorine; 10 p.p.t.v. bromine). Panel **a** (right-hand axis) also shows the seasonal NH chemical ozone depletion diagnosed directly from the standard model run. For all model simulations, the mixing ratio of O₃ in the troposphere (below ~12 km) was assumed to be 50 p.p.b.v. in the NH and 40 p.p.b.v. in the SH.

(March $\langle O_3 \rangle_{63-90}$) decreased from values near 470 dobson units (DU) in the early 1970s to values as low as 355 DU in 1997. They chose to analyse this region as it typically contains the polar vortex, which is often decaying by this time. We note, however, that the size of the vortex in March is very variable and, even with a conservative (large) definition of the vortex extent, this diagnostic will include mid-latitude air. For example, the size of the polar vortex was around 50% of this area in March 1992 and 80% in March 1997 (although the large vortex was not zonally symmetric). Figure 1a shows the monthly $\langle O_3 \rangle_{63-90}$ observations of Newman *et al.*¹ for the Arctic (March) and Antarctic (October) from 1990 extended to the end of 1998. The Antarctic $\langle O_3 \rangle_{63-90}$ values are small and show little interannual variability. This is a result of the large chemical removal of ozone within the large, stable Antarctic polar vortex which occurs

regularly every year. In contrast, the Arctic March $\langle O_3 \rangle_{63-90}$ values are generally much larger and show more interannual variability. In particular, the lowest March value in 1997 of around 355 DU was followed by a much larger value in 1998 of around 428 DU. In fact, the modelled $\langle O_3 \rangle_{63-90}$ in March 1997 indicates only a net 5 DU increase since November 1996.

The standard 3D CTM integration captures well the observed interannual variability in $\langle O_3 \rangle_{63-90}$ in the Arctic region, to within ~ 20 DU. The CTM reproduces all of the relative minima and maxima and, in particular, the strong minimum in 1997, followed by the large increase in 1998. (The CTM also reproduces synoptic-scale features which determine the time evolution of column ozone at specific sites related to the motion of the polar vortex¹⁹.) Given the excellent agreement, we have diagnosed the relative roles of chemistry and transport within the model. We have quantified the modelled rapid winter/spring chemical O_3 loss by two methods. First, we used a passive O_3 tracer to diagnose the total chemical loss each winter. On 15 December of each year this tracer was initialized with the model O_3 field. Second, we repeated the model experiment without heterogeneous reactions on solid PSCs or direct chlorine-activating reactions on cold liquid aerosols. The first method gives a direct quantification of the chemical O_3 loss in each winter; the second method illustrates the effect of polar processing in all seasons throughout the integration. Over the latitude range $63-90^\circ$ N, the average column chemical loss is around 30–40 DU ($\sim 50\%$ larger than the difference between the two 3D model experiments as the

different ozone fields feed back into the transport via calculated descent rates). This mean chemical loss reflects the relatively small area of the Arctic polar vortex relative to the area poleward of 63° N (the maximum local mean model column loss in March 1997 is around 80 DU, similar to other calculations for the mid-March loss of that year²¹) and that distortions to the polar vortex can cause O_3 -depleted vortex air to extend equatorward of 63° N. When presented as this monthly average over the area $43-90^\circ$ N, which includes air both inside and outside the vortex, the modelled interannual variability in the rapid vortex ozone depletion in March is relatively small, although the cold winters of 1995/96 and 1996/97 do show the largest chemical loss. While we acknowledge that the model simulation (run at low resolution) may underestimate the magnitude of the observed chemical loss—as shown in the Antarctic comparison in Fig. 1a and indicated by other model studies^{11,22}—the interannual variability in the vortex chemical loss is much less than the overall observed and modelled variability in column O_3 north of 63° N.

Figure 1b shows the $\langle O_3 \rangle_{63-90}$ from the model simulations at the start of the winters (November in the Northern Hemisphere, NH; May in the Southern Hemisphere, SH). The model predicts similar mean ozone values in the two polar regions, with only small interannual variation (for example, a standard deviation (s.d.) of 8 DU in the NH). This is supported by our analysis of Nimbus 7 TOMS data for November and May averaged over the sunlit area of the polar regions (H. Teyss  re, personal communication). This

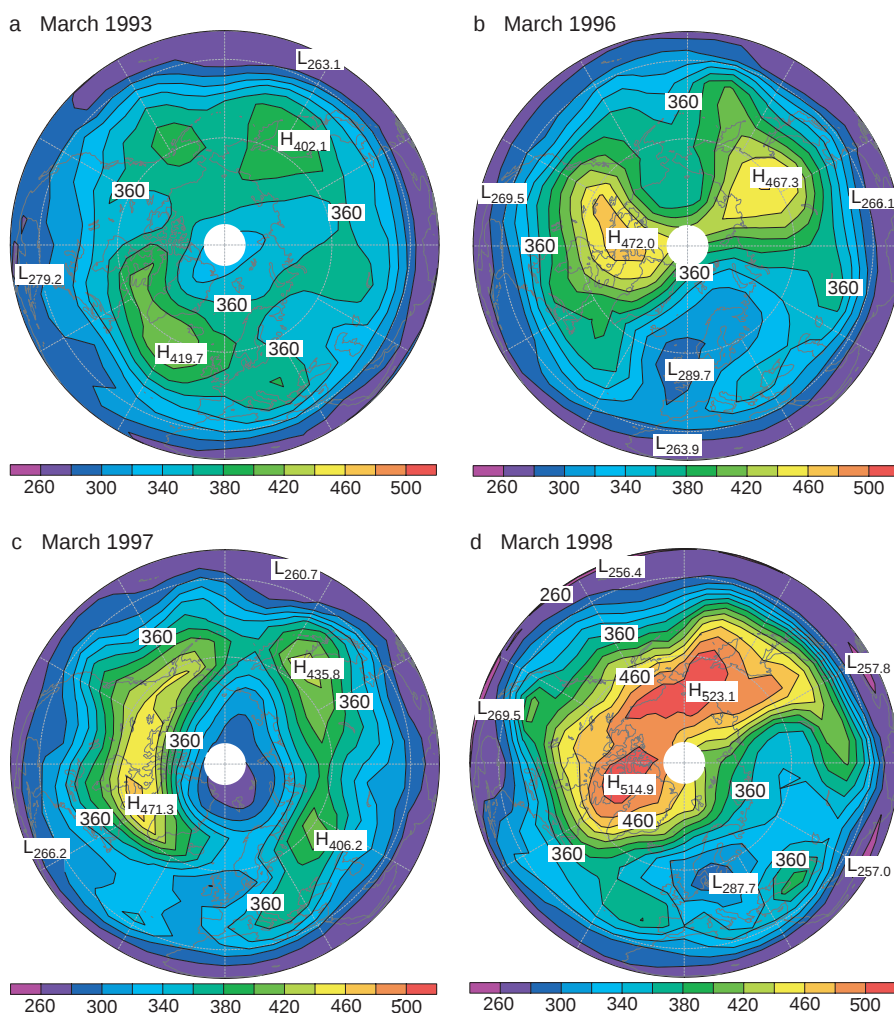


Figure 2 Northern Hemisphere March average column ozone (DU). Results are from the standard model run for 1993 (a), 1996 (b), 1997 (c) and 1998 (d). Contour

interval is 20 DU. The symbols H and L indicate relative maxima and minima, respectively.

suggests that all NH winters start with similar values of $\langle O_3 \rangle_{63-90}$ and that the interannual differences observed in March are determined by transport (and to a lesser degree, chemistry) during the winter. Comparison of Fig. 1a and b emphasizes that over the winter the SH experiences a strong net reduction due to the domination of chemical loss, while the NH experiences a strong increase due to transport which is partly offset by chemical loss. Having already diagnosed the chemical loss, and having shown that the modelled November means are similar from year to year, we can argue that the interannual differences in March are due to differences in the wintertime transport, which affect the size and stability of the vortex. Descent, and transport of ozone, is weakest when the vortex is strong and undisturbed, which are also the conditions that lead to cold temperatures. In contrast, a warm, disturbed vortex will result in stronger diabatic descent, and a larger increase in ozone. Further comparison of Fig. 1a and b shows that the dynamical build-up of O_3 during winter 1997/98 was particularly strong. The Arctic mean O_3 column for November 1997 was slightly lower than other years but the March 1998 column was much larger. Long-lived model tracers (such as total inorganic chlorine) indicate that the modelled wintertime high-latitude descent was indeed strongest in 1997/98 (not shown).

Figure 2 shows the March monthly mean O_3 for selected winters from the standard model run. The values of $\langle O_3 \rangle_{63-90}$ generally combine low column values associated with the mean position of the polar vortex and much higher values in the extra-vortex region (see, for example, the 1990s data of Figure 2 in ref. 1). In 1998 the modelled values in this larger extra-vortex region were much higher than previous model years.

The meteorological analyses used to force the CTM are only available from October 1991, so we cannot study directly the observed ozone variation throughout the 1970s and 1980s. However, we have investigated the effect of the increased atmospheric halogen loading by repeating the first 2 years of the standard model run with reduced halogen loadings typical of the early 1970s²³ (Cl_2 , 1.35 p.p.b.v.; Br_2 , 10 p.p.t.v.). This reduced halogen loading yields Arctic $\langle O_3 \rangle_{63-90}$ values which are higher by 8% in November and 14% in March. This March difference is similar in magnitude to the reduction observed by Newman *et al.*¹ between the early 1970s and early 1990s. Comparing Figs 3 and 2a shows that under the low-halogen conditions the calculated March ozone column is larger throughout the polar region; not just in the region of the polar vortex, but also throughout the 'collar' region. These model results suggest that, although interannual differences in wintertime halogen-catalysed ozone

loss is a minor contributor to the observed $\langle O_3 \rangle_{63-90}$ variation in the 1990s (that is, the s.d. of chemical loss is 8 DU compared to an s.d. of the March $\langle O_3 \rangle_{63-90}$ of 21 DU), the increase in halogens has nevertheless caused a reduction in Arctic $\langle O_3 \rangle_{63-90}$. This loss is due to increased destruction throughout the stratosphere throughout the year, which is not related to PSCs. Figure 4 shows differences in the November and March zonal-mean ozone fields. The increased halogens have caused a reduction of around 1 p.p.m.v. (up to 20%) in the upper stratosphere, and smaller mixing-ratio reductions—which will contribute to the column decrease—throughout the lower stratosphere which are not linked to polar processes. Diagnosis of the model shows that the main catalytic cycles responsible for this decrease involve $ClO^+ + O^+$ in the upper stratosphere, and cycles involving $ClO^+ + BrO^+$ or HO_2^+ in the lower stratosphere, with $ClO^+ + ClO^+$ important in the polar lower stratosphere.

Overall, our multiannual 3D model simulations have quantified many features of the observed interannual variations in Arctic ozone. Although PSC-related halogen chemistry certainly destroys O_3 within the polar vortex, the model shows that the mean spring-time abundance of ozone in the entire polar region (north of $63^\circ N$) is mainly controlled by dynamics acting over the winter (the mean dynamical increase in $\langle O_3 \rangle_{63-90}$ from November to March in our 7-year simulation was 80 DU, partially offset by a mean chemical loss of 38 DU). This dynamical build-up shows strong interannual variability which controls the observed interannual variability in the mean column ozone. Over this large area, the signal of rapid halogen-catalysed loss inside the vortex (which can be as small as

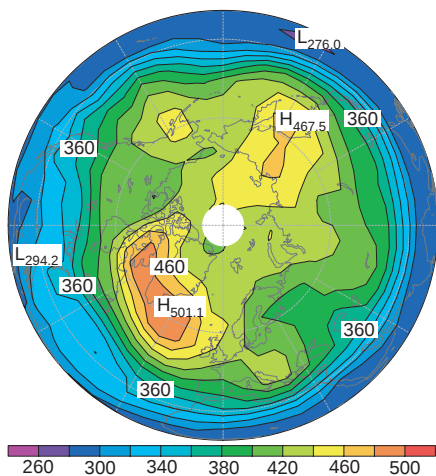


Figure 3 Northern Hemisphere average column ozone (DU) for March 1993. Results are from the model run with low halogen loading. Contour interval is 20 DU. The symbols H and L indicate relative maxima and minima, respectively.

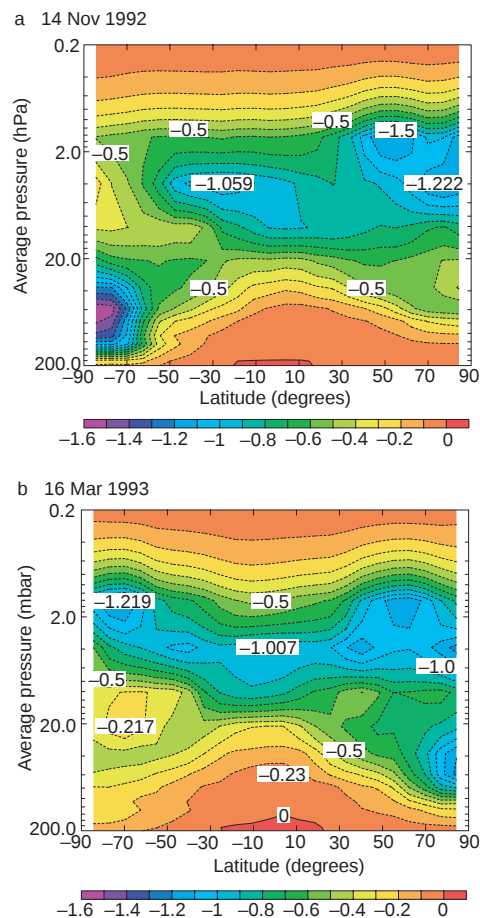


Figure 4 Difference in zonal mean O_3 mixing ratio (p.p.m.v.). Shown is the difference between the low-halogen experiment and the standard model run. **a**, 14 November; **b**, 16 March.

50% of this area in March) is reduced. However, the increases in stratospheric halogen loading due to anthropogenic emissions has contributed significantly to the springtime decrease since the 1970s. The observed decadal decrease in column O_3 values in March north of $63^\circ N$ may contain a large ($>50\%$) contribution from slow, year-round, halogen-catalysed depletion.

These results have implications for future levels of ozone in the Arctic region. Although the amount of PSC-induced chemical depletion within the vortex may be enhanced due to stratospheric cooling²⁴, the main factor determining the average ozone column to date in a given winter is dynamics. This transport is affected by interannual variability, but may also be subject to a trend—for example, due to stratospheric O_3 decrease or other climate-related effects. Possible future circulation changes (for example, those that make the Arctic vortex more Antarctic-like) could significantly change mean wintertime ozone levels by transport alone. Also, while Shindell *et al.*²⁴ pointed out the potential feedback between stratospheric cooling which may delay the recovery of Arctic ozone as halogen levels decrease, the contribution of halogens to the decadal decrease in ozone in the larger extra-vortex region may be expected to follow the halogen loading more closely, unless cooling extends PSC activation outside the vortex. □

Received 26 October 1998; accepted 11 June 1999.

1. Newman, P. A. *et al.* Anomalous low ozone over the Arctic. *Geophys. Res. Lett.* **24**, 2689–2692 (1997).
2. Fioletov, V. E. *et al.* Long-term ozone decline over the Canadian Arctic to early 1997 from ground-based and balloon observations. *Geophys. Res. Lett.* **24**, 2705–2708 (1997).
3. Rex, M. *et al.* Prolonged stratospheric ozone loss in the 1995–96 Arctic winter. *Nature* **389**, 835–838 (1997).
4. Müller, R. *et al.* Chlorine activation and ozone depletion in the Arctic vortex: Observations by the Halogen Occultation Experiment on the Upper Atmosphere Research Satellite. *J. Geophys. Res.* **101**, 12531–12554 (1996).
5. Müller, R. *et al.* Severe chemical ozone loss in the Arctic during the winter 1995–96. *Nature* **389**, 709–712 (1997).
6. Müller, R. *et al.* HALOE observations of the vertical structure of chemical ozone depletion in the Arctic vortex during winter and early spring 1996–1997. *Geophys. Res. Lett.* **24**, 2717–2720 (1997).
7. Manney, G. L. *et al.* Arctic ozone depletion observed by UARS MLS during the 1994–95 winter. *Geophys. Res. Lett.* **23**, 85–88 (1996).
8. Manney, G. L., Santee, M. L., Froidevaux, L., Waters, J. W. & Zurek, R. W. Polar vortex conditions during the 1995–96 Arctic winter: Meteorology and MLS ozone. *Geophys. Res. Lett.* **23**, 3203–3206 (1996).
9. Manney, G. L., Froidevaux, L., Santee, M. L., Zurek, R. W. & Waters, J. W. MLS observations of Arctic ozone loss in 1996–97. *Geophys. Res. Lett.* **24**, 2697–2700 (1997).
10. Sinnhuber, B. M., Langer, S., Klein, U., Raffalski, U. & Kunzi, K. Ground based millimeter-wave observations of Arctic ozone depletion during winter and spring of 1996/97. *Geophys. Res. Lett.* **25**, 3327–3330 (1997).
11. Hansen, G., Svenøe, T., Chipperfield, M. P., Dahlback, A. & Hoppe, U. P. Evidence of substantial ozone depletion in winter 1995/96 over northern Norway. *Geophys. Res. Lett.* **24**, 799–802 (1997).
12. Zurek, R. W., Manney, G. L., Miller, A. J., Gelman, M. E. & Nagatani, R. M. Interannual variability of the north polar vortex in the lower stratosphere during the UARS mission. *Geophys. Res. Lett.* **23**, 289–292 (1996).
13. Dobson, G. M. B. Observations of the amount of ozone in the earth's atmosphere and its relation to other geophysical conditions. *Proc. R. Soc. Lond. A* **129**, 411–433 (1930).
14. Hood, L. L., McCormack, J. P. & Labitzke, K. An investigation of dynamical contributions to midlatitude ozone trends in winter. *J. Geophys. Res.* **102**, 13079–13093 (1997).
15. Steinbrecht, W., Claude, H., Kohler, U. & Hoinka, K. P. Correlations between tropopause height and total ozone: Implications for long-term changes. *J. Geophys. Res.* **103**, 19183–19192 (1998).
16. Rood, R. B., Allen, D. J., Baker, W. E., Lamich, D. J. & Kaye, J. A. The use of assimilated stratospheric data in constituent transport calculations. *J. Atmos. Sci.* **46**, 687–701 (1989).
17. Lefevre, F., Brasseur, G. P., Folkins, I., Smith, A. K. & Simon, P. Chemistry of the 1991–1992 stratospheric winter: Three-dimensional model simulations. *J. Geophys. Res.* **99**, 8183–8195 (1994).
18. Chipperfield, M. P. *et al.* Analysis of UARS data in the southern polar vortex in September 1992 using a chemical transport model. *J. Geophys. Res.* **101**, 18861–18881 (1996).
19. Chipperfield, M. P. Multiannual simulations with a 3D chemical transport model. *J. Geophys. Res.* **104**, 1781–1805 (1999).
20. Swinbank, R. & O'Neill, A. A stratosphere-troposphere data assimilation system. *Mon. Weath. Rev.* **122**, 686–702 (1994).
21. Lefevre, F., Figarol, F., Carslaw, K. S. & Peter, T. The 1997 Arctic ozone depletion quantified from three-dimensional model simulations. *Geophys. Res. Lett.* **25**, 2425–2428 (1998).
22. Becker, G., Müller, R., McKenna, D. S., Rex, M. & Carslaw, K. S. Ozone loss rates in the Arctic stratosphere in the winter 1991/1992: Model calculations compared with Match results. *Geophys. Res. Lett.* **25**, 4325–4329 (1998).
23. World Meteorological Organization *Scientific Assessment of Ozone Depletion: 1994* (Rep. No. 7, Global Ozone Research and Monitoring Project, Geneva, 1995).
24. Shindell, D. T., Rind, D. & Lonerger, P. Increased polar stratospheric ozone losses and delayed eventual recovery owing to increasing greenhouse-gas concentrations. *Nature* **392**, 589–592 (1998).

Acknowledgements. We thank P. Newman for information about the 1998 TOMS data, H. Teyssède for TOMS calculations, and J. A. Pyle for help and support. This work was supported by the UK Natural Environment Research Council.

Correspondence and requests for materials should be addressed to M.P.C. (e-mail: martyn@lec.leeds.ac.uk).

2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis

Roger E. Summons*, Linda L. Jahnke, Janet M. Hope* & Graham A. Logan*

* Australian Geological Survey Organisation, GPO Box 378, Canberra, ACT 2601, Australia

† Exobiology Biology Branch, NASA Ames Research Center, Moffett Field, California 94035, USA

Oxygenic photosynthesis is widely accepted as the most important bioenergetic process happening in Earth's surface environment¹. It is thought to have evolved within the cyanobacterial lineage, but it has been difficult to determine when it began. Evidence based on the occurrence and appearance of stromatolites² and microfossils³ indicates that phototrophy occurred as long ago as 3,465 Myr although no definite physiological inferences can be made from these objects. Carbon isotopes and other geological phenomena^{4,5} provide clues but are also equivocal. Biomarkers are potentially useful because the three domains of extant life—Bacteria, Archaea and Eukarya—have signature membrane lipids with recalcitrant carbon skeletons. These lipids turn into hydrocarbons in sediments and can be found wherever the record is sufficiently well preserved. Here we show that 2-methylbacteriohopanepolyols occur in a high proportion of cultured cyanobacteria and cyanobacterial mats. Their 2-methylhopane hydrocarbon derivatives are abundant in organic-rich sediments as old as 2,500 Myr. These biomarkers may help constrain the age of the oldest cyanobacteria and the advent of oxygenic photosynthesis. They could also be used to quantify the ecological importance of cyanobacteria through geological time.

In the Bacteria, bacteriohopanepolyols (BHP) are amphiphilic membrane biochemicals⁶ that serve a regulating and rigidifying function similar to that of sterols in the Eukarya. The hydrocarbon skeletons of BHP are extremely refractory and resist biodegradation to become incorporated into kerogen or bound into sulphur-linked macromolecules. They also survive high geothermal gradients accompanying hydrocarbon generation and ultimately appear in petroleum as defunctionalized and stereochemically modified⁷ 'geohopanes'. Although many aspects of the preservation of BHP are well understood, there remains uncertainty about the sources of BHP and environmental controls on their abundance and distribution. This study assesses the role of cyanobacteria as a source of fossil hopanoids.

Table 1 summarizes the hopanoid contents of cultured cyanobacteria and mat communities that we analysed, along with data extracted from the literature. The high polarity and complexity of the hydrophilic side chains of BHP⁸ precludes a generalized gas chromatography–mass spectrometry (GC–MS) analysis that detects all hopanoids. Accordingly, we applied the methodology of ref. 9 (Fig. 1) in our survey of the occurrence of hopane skeletons in cultures and environmental samples. Structural variation in the hydrophobic pentacyclic triterpane skeleton of BHP appears to be very limited. Unsaturation ($\Delta 6$ and/or $\Delta 11$) is rare, but supplementary methyl substituents at the 2α , 2β and 3β positions of ring-A have been reported often and may indicate metabolism in some bacteria. Methylation at C3 has been reported for a variety of methanotrophic, methylotrophic and acetic bacteria^{9,10}. Methylation at C2 is common in cyanobacteria^{9,11}, as shown in Table 1, although it is not exclusive to this group. Of those other bacteria that have been studied, 2-methylhopanoids have also been found in pink pigmented facultative methylotrophs (PPFMs) related to *Methylobacterium organophilum*¹² and some nitrogen-fixing bacteria

Table 1 Abundances ($\mu\text{g per g dry wt}$) of hopanols analysed in cultured cyanobacteria, a prochlorophyte and various environmental samples

Cyanobacterial cultures and environmental samples analysed in this study	C31 5	2-MeC31 6	C32 7	2-MeC32 8
<i>Phormidium luridum</i> UTEX 426 (CCAP 1462/2)			540	74
<i>Chlorogloeopsis fritschii</i> ATCC 27193			65	7
<i>Synechococcus lividus</i> ATCC 27180			155	95
<i>Synechococcus</i> ATCC 29534 (<i>Aphanothece halophitica</i>)			20	30
<i>Synechococcus</i> ATCC 27144 (<i>Anacystis nidulans</i>)	39	63	25	23
<i>Gloeobacter violaceus</i> ATCC 29082			450	650
<i>Anabaena cylindrica</i> ATCC 27899			68	
<i>Fischerella</i> sp. ATCC 29538			2	
<i>Phormidium</i> sp. 'OSS4' Octopus Spring, YNP	179	258	<1	23
<i>Phormidium</i> sp. 'RCG' Rabbit Creek, YNP			161	
<i>Phormidium</i> sp. 'RCO' Rabbit Creek, YNP			<1	
<i>Phormidium</i> sp. 'OSS3' Octopus Spring, YNP			45	
<i>Chlorogloeopsis</i> sp. 'LA' Norris Geyser Basin, YNP	286		374	
<i>Calothrix</i> sp. 'TSB' Twin Springs, YNP			255	
<i>Oscillatoria limnetica</i> (Y. Cohen), Solar Lake, Israel			40	
<i>Oscillatoria</i> sp. 'GN' (E. Munoz), Guerrero Negro, Mexico			<1	
<i>Phormidium</i> sp. 'GN' (E. Munoz), Guerrero Negro, Mexico			<1	
<i>Oscillatoria</i> sp. (L. Stal) Mellum Island, Germany			27	
<i>Oscillatoria</i> sp. 10mfx (J. Bauld) Shark Bay, Australia			18	
Octopus Spring <i>Synechococcus-Chloroflexus</i> 60 °C mat*	14	20	23	
Octopus Spring <i>Synechococcus-Chloroflexus</i> 55 °C mat* ¹⁶	1	1	8	
Octopus Spring <i>Phormidium-Chloroflexus</i> 45 °C mat*	61	64	90	29
Fountain Paint Pot <i>Phormidium-Chloroflexus</i> 45 °C mat*	9	30	24	13
Rabbit Creek <i>Phormidium-Chloroflexus</i> 40 °C mat*	2	32	24	7
Shark Bay 'Smooth' Mat (J88/1 predom. <i>Microcoleus</i> sp.)†			1.2	
Shark Bay 'Tufted' Mat (J88/9 predom. <i>Lyngbya</i> sp.)†			0.7	
Shark Bay 'Smooth' Mat (J88/10 predom. <i>Microcoleus</i> sp.)†			0.7	
Shark Bay 'Tufted' Mat (J88/14 predom. <i>Lyngbya</i> sp.)†			20	
Shark Bay 'Pustular' Mat (J88/17 predom. <i>Entophysalis</i> sp.)†			1.0	
Data extracted from literature reports				
<i>Prochlorothrix hollandica</i> whole cells ¹¹			60	90
<i>Prochlorothrix hollandica</i> cell wall ¹¹			400	600
<i>Prochlorothrix hollandica</i> thylakoids ¹¹			940	1410
<i>Nostoc</i> sp. ATCC 27985† ⁸			250	
<i>Nostoc muscorum</i> B-1453-12b† ²⁶			3x	7x
<i>Nostoc</i> sp. PCC 6720 ²⁷	55	220	66	160
<i>Scytonema</i> sp. ATCC 29171† ⁸			100	
<i>Synechococcus</i> sp. PCC 6907 ²⁸			135	72
<i>Calothrix anomala</i> ¹⁵	15		<1	400
<i>Schizothrix</i> sp. ¹⁵	2		130	<1
<i>Anacystis montana</i> ¹⁵	<1		44	180
<i>Oscillatoria rubescens</i> ¹⁵	<1		430	<1
<i>Gloeobacter</i> sp. ¹⁵	<1		<1	2700
Various species (9)‡			3-140	
Various species (3)‡ ⁸	<1		<1	

* Indicates relative amounts of bacteriohopanol in mats and are for top photic zone (~1 mm for *Synechococcus* and up to 1 cm for *Phormidium*) and do not account for high levels of silica associated with dry weight.

† Indicates data do not account for high levels of CaCO_3 associated with dry weight.

|| Indicates an x-methylhopanol was detected but the identity and proportion are unspecified.

¶ Indicates only compound ratios and not absolute abundances available.

Indicates that a 2-methyl- C_{30} hopanol precursor was also reported.

‡ Indicates that the original 1984 methodology was used and this may have underestimated BHP contents (M. Rohmer, personal communication).

of the genera *Azotobacter* and *Beijerinckia*¹³. However, the predominant 2-methylhopanoids in these organisms are diplopterols and diploptenes, which are C_{31} alcohols and hydrocarbons. Only trace amounts of 2-methyl-BHP have been identified in *M. organophilum*¹⁴. This has significant implications for their preservation potential, because diplopterols readily dehydrate to diploptenes, which have very limited capacity to be incorporated into kerogen. BHP and 2-methyl-BHP, on the other hand, have multiple functional groups through which they can be incorporated into the kerogen matrix and preferentially preserved. Although 2-methyl-BHP may be found in unstudied taxa in the future, the cyanobacterial lineage currently comprises the only known quantitatively significant source^{9,11,15,16} with a conduit to selective preservation.

The C_{32} hopanol (alcohol precursor of type 7 in Fig. 1) is by far the most common BHP derivative encountered in cyanobacteria. Its BHP precursors are probably the source of the C_{32} carboxylic acid and C_{31} hydrocarbon that are the most widespread geohopanooids in recent sediments. Around 43% of the listed samples contain 2-methyl analogues (6 and 8). Four out of five mats from the Yellowstone National Park hydrothermal environment contained significant 2-methyl-BHP, although none of five samples from the hypersaline environments of Shark Bay did. Although the present sample set gives limited coverage of modern environments, species

that have been cultured include one prochlorophyte and span a taxonomically diverse spectrum of cyanobacteria. The Yellowstone *Phormidium* conophyton mats that we studied all had abundant 2-methyl-BHP. These are the possible modern analogues for the conophyton stromatolites that have a sedimentary record reaching back 2,700 Myr (ref. 2). However, note that biomarker probes for the antiquity of cyanobacteria are unidirectional, as 2-methyl-BHP are not found in all cultured species nor all mats.

The sedimentary conversion of BHP to the $\text{C}_{27}\text{-C}_{35}$ geohopane series (3 in Fig. 1) has been extensively studied in connection with the application of these biomarkers to petroleum exploration⁷. The burial fate of 2-methyl-BHP is virtually identical to that of BHP except that the 2 β -methyl configuration of the original biochemicals (2 in Fig. 1) is progressively converted to the more thermodynamically stable 2 α -methyl configuration (4 in Fig. 1) in the 2-methyl geohopane series¹⁷. Despite this, we need to be cautious in how we validate fossil hydrocarbons. Establishing that they are indigenous to the host rock requires close attention to a variety of factors, as these compounds are mobile. Low levels of metamorphism, high contents of associated kerogen, covariance of the $\delta^{13}\text{C}$ of kerogen and bitumen, remoteness from younger oil-prone rocks and distinctive compound assemblages indicate syngeneity, especially when there is convergence of multiple lines of evidence for a particular sample. To the best of our knowledge, all

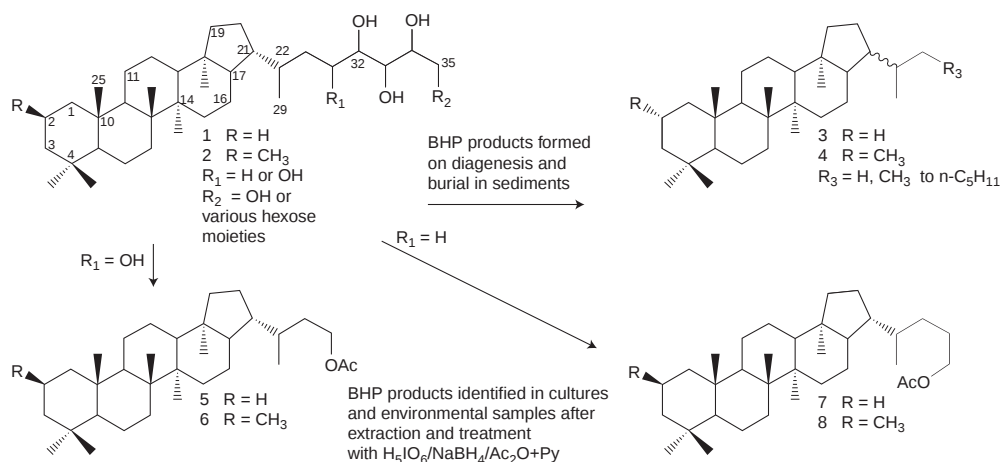


Figure 1 Structures of bacteriohopanepolyols (1, 2) typical of those found in cyanobacteria and their geohopane analogues recognized in sediments and

petroleum (3, 4). BHP from cultures and environmental samples were chemically reduced to hopanol acetates of types 5–8 and analysed by GC–MS.

the bitumens listed in Table 2 are indigenous to their host formations, whereas all the oils are derived from specific source rocks of the designated ages.

The data shown in Table 2 confirm earlier observations^{17,18} that higher relative abundances of 2-methylhopanes, expressed here as a 2-methylhopane index, are a distinctive feature of oils and bitumens derived from carbonate rocks. Moreover, there are high relative abundances of 2-methylhopanes in Proterozoic (>550 Myr) sediments of all lithologies compared to their Phanerozoic (<550 Myr) counterparts. Our data indicate that burial temperature strongly affects the 2-methylhopane index, because all immature samples analysed had low values. This indicates that generation requires cracking from chemically-bound precursors and independently supports their origin from BHP rather than diplopterols or diploptenes. The data for mature samples confirm previous observations that oils from carbonate source rocks have the highest 2 α -methylhopane indices¹⁷. Such sediments frequently form in warm, low-

latitude, shallow-water continental shelves, coastal lagoons, and other restricted marine settings. They also form in playa and saline lakes, as exemplified by the samples from the Early Cambrian Observatory Hill Formation, which had the highest 2-methylhopane indices we have encountered so far. Modern settings of this type are noted for the development of taxonomically diverse cyanobacterial mats^{19,20} and moderate-to-high rates of organic matter accumulation²⁰. It is likely that this has been the case for at least the past 2,500 Myr. Given the occurrence of 2-methyl-BHP in a wide variety of cyanobacterial taxa and cyanobacterial mats, a high abundance of 2-methylhopanes is expected in carbonate-derived oils of any age. Organic-rich shales, on the other hand, form in more open and distal marine settings including shelves, slopes and deltas, prone to clastic sediment inputs. In modern and Phanerozoic settings such as these, algae are the prime autotrophs with sedimentary carbon supplemented with inputs from allochthonous terrestrial organic matter. Planktonic and picoplanktonic cyanobacteria may be responsible for significant primary productivity in the modern ocean and our data indicate that they would have been more significant in times past, particularly in the Proterozoic²¹. Oils and bitumens from Proterozoic shales have higher 2-methylhopane indices (5–18.7%) than the mature Phanerozoic shale samples that we analysed (0.3–9%).

Given that 2-methyl-BHP are prominent lipids in cyanobacteria and cyanobacteria-dominated environmental samples, it is evident that 2-methyl-BHP and their derivative 2 α -methylhopanes also comprise a molecular record for oxygenic photosynthesis, as this is their preferred physiology. The high abundance of derivative 2 α -methylhopanes in bitumens from the 2,500 Myr Mt McRae Shale of the Hamersley basin provides direct evidence for oxygenic photosynthesis at this time and supports microfossil and other geological evidence that the process predates this. The Hamersley basin samples are also significant for another reason. Their sediment extracts extend the molecular biomarker record to the base of the Palaeoproterozoic, and are 700 Myr older than the oldest known biomarkers²². The Mt McRae Shale meets most of the important syngeneity criteria, especially with respect to having high organic carbon contents (0.5–6.9%) and hydrocarbon stereoisomer patterns that correlate well with other estimates of their thermal history ($R_o \approx 1.5$ –1.9). The Pilbara craton of Western Australia, location of the Hamersley basin, is also remote from any known Phanerozoic petroleum sources. Although it is difficult to find suitably preserved Archaean (>2,500 Myr) samples, recent work on fluorescing, fluid-bearing inclusions in quartz crystals in sandstones of this age, and older, indicate that some hydrocarbons generated at this time may have been preserved²³ and are likely to

Table 2 2-Methylhopane indices for well characterized oils and sedimentary bitumens

Formation	Lithology	Age (Myr)	Index
Mt. McRae Shale (3)	Shale	~2,500	12.5
Wollogorang Fm. (2)	Shale	~1,730	12.0
Barney Creek Fm. (6)	Shale	~1,640	8.5
Balbirini Fm.	Shale	~1,600	9.0
Yalco Fm. (1)	Shale	~1,485	14.7
McMinn Fm. (1)	Shale	~1,400	5.8
Nonesuch Fm. (2)	Shale	~1,000	4.7
Kwagunt Fm. (2)	Marl	~850	10.9
Kwagunt Fm. (1)	Shale	~850	18.7
Bitter Springs Fm (3)	Marl	~830	6.6
Draken Fm. (2)	Carbonate	~720	8.8
Ungoolya Fm. (4)	Shale	~560	6.7
Pertatataka Fm. (3)	Shale	~560	4.9
Shuram Fm. (1)	Marl	~560	12.7
Athel Fm. (1)	Marl	~550	11.2
Huqf Group oil (2)	Carbonate	~550	10.9
Observatory Hills Fm. (4)	Carb./evap.	530–540	32.4
Ouldburra Fm. (4)	Carbonate	530–540	9.9
Inca Fm.* (1)	Shale	505–510	2.4
Horn Valley Fm. (1)	Shale	460–485	5.3
Milligans Fm. (2)	Shale	330–345	9.4
Phosphoria Fm. oil (1)	Shale	270–300	0.5
Cooper basin oils (4)	Coal	250–270	4.0
Bowen basin oils (12)	Coal	250–270	8.9
Vulcan sub-basin oil (2)	Shale	150–160	4.5
Barrow sub-basin oil (2)	Shale	150–160	4.5
North Sea oil (1)	Shale	150–160	2.5
Julia Ck. Fm.* (1)	Shale	97–108	0.6
Taranaki basin (1)	Shale	61–65	0.3
Monterey Fm.* (2)	Carbonate	17–24	5.0
Monterey Fm.* (1)	Shale	6–8	0.9

* Samples above the oil window (see Methods).

hold a valid chemical fossil record²⁴.

Cyanobacterial BHP and 2-methyl-BHP may have oceanographic and palaeo-oceanographic applications as well, particularly with respect to evaluation of cyanobacterial primary productivity and their importance for the marine nitrogen and carbon cycles²⁵. As biomarkers such as BHP and chlorophylls carry ¹³C and ¹⁵N signatures¹⁶, their usefulness as tracers in modern aquatic and marine environments is significantly broadened. □

Methods

Total lipid extracts from cultures and environmental samples were analysed using a procedure modified and improved after ref. 8. Periodic acid oxidation followed by NaBH₄ reduction converted BHP to simpler hopanols amenable to purification by thin-layer chromatography and GC-MS analysis as acetate derivatives. BHP lacking a gem-diol function evade detection. The total number of culture samples with suitable data was 42 and, of these, only 6 failed to yield detectable hopanol. The data in Table 1 are from this study or directly extracted or calculated from published quantitative data.

The data for fossil hopanoids shown in Table 2 was derived from GC-MS-MS analyses⁵ and based on the *m/z* 412 → 191 transition for αβ-hopane (3 in Fig. 1, R₃ = H) and *m/z* 426 → 205 for its 2α-methyl analogue (4, R₃ = H). These are expressed as the 2-methylhopane index = % 4/4+3. Accompanying C₂₈–C₃₆ homohopanes (R = CH₃, R₃ = CH₃ to C₅H₇) were examined using the *m/z* 205 ion chromatograms and confirmed the relationships expressed here for the C₃₀ and 2α-methyl C₃₁ species. Age and lithology assignments for Phanerozoic samples were derived from AGSO databases. Assignments for Proterozoic sediments are primarily based on ref. 5, updated where possible. Numbers in parentheses indicate (*n*) samples of each rock unit. All samples are mature for oil generation except those marked with asterisks, which are above the oil window. Low 2α-methylhopane indices in these samples indicates incomplete release by thermal cracking of C₃₅ and 2-Me C₃₆ moieties bound in kerogen.

Received 23 March; accepted 11 June 1999.

- Blankenship, R. E. & Hartman, H. The origin and evolution of oxygenic photosynthesis. *Trends Biochem. Sci.* **23**, 94–97 (1998).
- Walter, M. R. in *Early Life on Earth* (ed. Bengtson, S.) 270–286 (Columbia Univ. Press, New York, 1994).
- Schopf, J. W. Microfossils of the early Archean Apex Chert: New evidence of the antiquity of life. *Science* **260**, 640–646 (1993).
- Buick, R. The antiquity of oxygenic photosynthesis: Evidence from stromatolites in sulfate-deficient Archean lakes. *Science* **255**, 74–77 (1992).
- Schopf, J. W. & Klein, C. *The Proterozoic Biosphere. A Multidisciplinary Study* (Cambridge Univ. Press, 1992).
- Jürgens, U. J., Simonin, P. & Rohmer, M. Localisation and distribution of hopanoids in membrane systems of the cyanobacterium *Synechocystis* PCC 6714. *FEMS Microbiol. Lett.* **92**, 285–288 (1992).
- Peters, K. E. & Moldowan, J. M. *The Biomarker Guide—Interpreting Molecular Fossils in Sediments and Petroleum* (Prentice-Hall, Englewood Cliffs, New Jersey, 1992).
- Rohmer, M. The biosynthesis of triterpenoids of the hopane series in the eubacteria: a mine of new enzyme reactions. *Pure Appl. Chem.* **65**, 1293–1298 (1993).
- Rohmer, M., Bouvier-Nave, P. & Ourisson, G. Distribution of hopanoid triterpenes in prokaryotes. *J. Gen. Microbiol.* **130**, 1137–1150 (1984).
- Zundel, M. & Rohmer, M. Prokaryotic triterpenoids 3. The biosynthesis of 2β-methylhopanoids and 3β-methylhopanoids of *Methylobacterium organophilum* and *Acetobacter pasteurianus* spp. *pasteurianus*. *Eur. J. Biochem.* **150**, 35–39 (1985).
- Simonin, P., Jürgens, U. J. & Rohmer, M. Bacterial triterpenoids of the hopane series from the prochlorophyte *Prochlorothrix hollandica* and their intracellular localisation. *Eur. J. Biochem.* **241**, 865–871 (1996).
- Knani, M., Corpe, W. A. & Rohmer, M. Bacterial hopanoids from pink-pigmented facultative methylotrophs and from green plant surfaces. *Microbiology* **140**, 2755–2759 (1994).
- Vilchèze, C., Llopiz, P., Neunlist, S., Poralla, K. & Rohmer, M. Prokaryotic triterpenoids: new hopanoids from the nitrogen-fixing bacteria *Azotobacter vinelandii*, *Beijerinckia indica* and *Beijerinckia mobilis*. *Microbiology* **140**, 2794–2795 (1994).
- Renoux, J.-M. & Rohmer, M. Prokaryotic triterpenoids. New bacteriohopane tetrol cyclitol ethers from methylotrophic bacterium *Methylobacterium organophilum*. *Eur. J. Biochem.* **151**, 405–410 (1985).
- Hermann, D. Des biohopanoides aux géohopanoïdes. Une approche de la formation des fossiles moléculaires de triterpénoïdes en série hopane. Thesis, Univ. Haute-Alsace (1995).
- Summons, R. E., Jahnke, L. L. & Simoneit, B. R. T. in *Evolution of Hydrothermal Ecosystems on Earth (and Mars?)*: Ciba Foundation Symposium 202 174–194 (Wiley, Chichester, 1996).
- Summons, R. E. & Jahnke, L. L. in *Biomarkers in Sediments and Petroleum* (eds Moldowan, J. M., Albrecht, P. & Philp, R. P.) 182–200 (Prentice Hall, Englewood Cliffs, New Jersey, 1992).
- Summons, R. E. & Walter, M. R. Molecular fossils and microfossils of prokaryotes and protists from Proterozoic sediments. *Am. J. Sci.* **A 290**, 212–244 (1990).
- Bauld, J. in *Microbial Mats: Stromatolites* (eds Cohen, Y., Castenholz, R. W. & Halverson, H. O.) 39–58 (Alan Liss, New York, 1984).
- Kenig, F. et al. Occurrence and origin of mono-, di-, and trimethylalkanes in modern and Holocene cyanobacterial mats from Abu Dhabi, United Arab Emirates. *Geochim. Cosmochim. Acta* **59**, 2999–3015 (1995).
- Golubic, S. in *Early Life on Earth: Nobel Symposium No 84* (ed. Bengtson, S.) 220–236 (Columbia Univ. Press, New York, 1994).

- Hayes, J. M., Summons, R. E., Strauss, H., Des Marais, D. J. & Lambert, I. B. in *The Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 81–133 (Cambridge Univ. Press, 1992).
- Buick, R., Rasmussen, B. & Krapez, B. Archean oil: evidence for extensive hydrocarbon generation and migration 2.5–3.5 Ga. *AAPG Bull.* **82**, 50–69 (1998).
- Price, L. C. Thermal stability of hydrocarbons in nature: limits, evidence, characteristics, and possible controls. *Geochim. Cosmochim. Acta* **57**, 3261–3280 (1993).
- Haug, G. H. et al. Glacial/interglacial variations in production and nitrogen fixation in the Cariaco Basin during the last 580 kyr. *Paleoceanography* **13**, 427–432 (1998).
- Bisseret, P., Zundel, M. & Rohmer, M. Prokaryotic triterpenoids: 2. 2β-methylhopanoids from *Methylobacterium organophilum* and *Nostoc muscorum*, a new series of prokaryotic triterpenoids. *Eur. J. Biochem.* **150**, 29–34 (1985).
- Zhao, N. et al. Structures of two bacteriohopanoids with acyclic pentol side-chains from the cyanobacterium *Nostoc* PCC 6720. *Tetrahedron* **52**, 2772–2778 (1996).
- Llopiz, P., Jürgens, U. J. & Rohmer, M. Prokaryotic triterpenoids: Bacteriohopanetetrol glyuronosides from the thermophilic cyanobacterium *Synechococcus* PCC6907. *FEMS Microbiol. Lett.* **140**, 199–202 (1996).

Acknowledgements. We thank D. Taylor for providing Hamersley basin samples and associated geological and maturity data. M. Rohmer, J. Hayes, K. Hinrichs, A. Knoll and M. Walter provided critical and constructive comments on drafts of this manuscript.

Correspondence and requests for material should be addressed to R.E.S. (e-mail: Roger.Summons@agso.gov.au).

Spatial scaling laws yield a synthetic theory of biodiversity

Mark E. Ritchie* & Han Olff†

* Department of Fisheries and Wildlife, Utah State University, Logan, Utah 84322-5210, USA

† Tropical Nature Conservation and Vertebrate Ecology Group, Department of Environmental Science, Wageningen Agricultural University, Bornsesteeg 69, 6708 PD Wageningen, The Netherlands

Ecologists still search for common principles that predict well-known responses of biological diversity to different factors^{1–4}. Such factors include the number of available niches in space^{5–7}, productivity^{8–10}, area¹⁰, species' body size^{11–14} and habitat fragmentation. Here we show that all these patterns can arise from simple constraints on how organisms acquire resources in space. We use spatial scaling laws to describe how species of different sizes find food in patches of varying size and resource concentration. We then derive a mathematical rule for the minimum similarity in size of species that share these resources. This packing rule yields a theory of species diversity that predicts relations between diversity and productivity more effectively than previous models^{8–10}. Size and diversity patterns for locally coexisting East African grazing mammals and North American savanna plants strongly support these predictions. The theory also predicts relations between diversity and area and between diversity and habitat fragmentation. Thus, spatial scaling laws provide potentially unifying first principles that may explain many important patterns of species diversity.

The search for a 'unified' theory of diversity^{1–5} has focused on the premise that more species can exist within a habitat whenever they can more finely divide up space and different-sized resource 'packages'. Such partitioning may be constrained by the different body sizes of species^{5,7,11,12–14}, but the mechanisms by which organism size, resource availability and spatial structure of habitats control species diversity remain unclear^{1,2,7,11,14}. Here we employ spatial scaling laws to describe how species with different body sizes find resources in space, and how limits to the similarity in body size between any two species predicts the potential number of species in a community.

Individual organisms must search within a space of suitable physical/chemical conditions (habitat) to find resources, which are often only available inside other material (food) (Fig. 1). Therefore, resources available to organisms are nested within food, and available food is nested within habitat. For example, insect herbivores move through suitable microclimates on terres-

trial plants (habitat) to eat plant tissue (food), which contains digestible carbohydrates (resources). Predatory fish search macrophyte-free areas of lakes (habitat) to eat invertebrates or smaller fish (food) that contain protein (resources). More imaginatively, terrestrial plants extend roots into rock-free soil (habitat) to take up soil solution (food) that contains nutrients (resources). Within a habitat, different species of similar trophic positions may harvest different sizes or types of food to obtain the same resources.

Distributions of habitat, food and resources often appear to be statistically self-similar (or self-affine) across ecologically relevant ranges of scales (3–4 orders of magnitude)^{15,16}. If so, their volume or area and spatial distribution can be described with fractal geometry, that is, simple scaling laws. The total amount of habitat within a landscape of extent x is hx^D , where D is the fractal dimension of the habitat and h is a prefactor¹⁷. Likewise, food patches occupy a volume mx^F and resources occupy a volume rx^Q . The fractal dimensions D , F and Q represent the degree to which habitat, food and resources fill space, and can vary from 0 (a single point) to 3 (a filled cube). However, the assumption that habitat, food and resources are nested distributions¹⁵ requires that $D \geq F \geq Q$. The prefactors h , m and r reflect the local density and contagion (lacunarity) of habitat, food and resources, respectively^{15–17}.

In a fractal environment, body size critically determines the abundance of food and resources that a species perceives^{17–19} (Fig. 1). Individuals sample a volume of space at a particular scale of resolution: the length w of the 'ruler' with which they perceive or sample the environment. This scale of resolution is presumably proportional to body size. If so, a species will subdivide its habitat

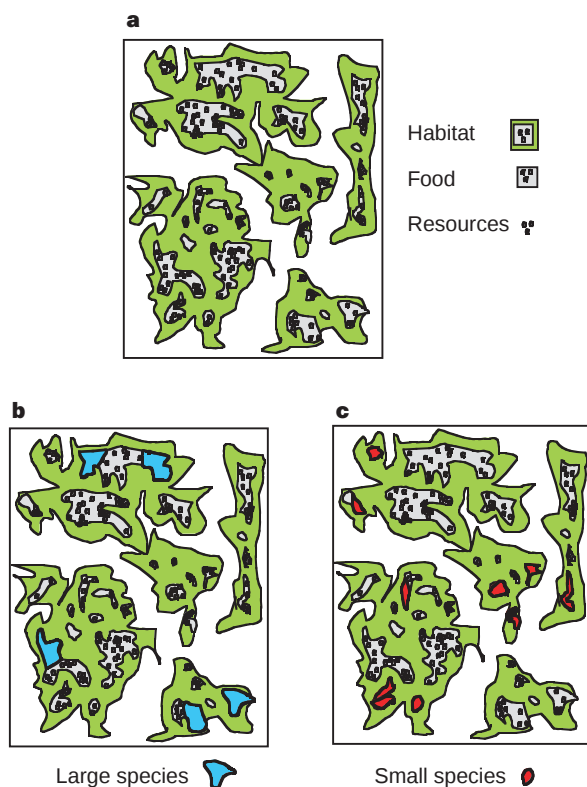


Figure 1 Hypothetical distribution of resources, food and habitat, as used by species of different size. **a**, Hypothetical space of extent x used by species of different size, including habitat, the food in which resources are contained and resources contained in food. Each element is fractal with the following mass fractal dimensions¹⁷: habitat: $D = 1.88$; food: $F = 1.53$; resource: $Q = 1.22$. Larger species (**b**) exclusively use large patches with low resource concentration, whereas smaller species (**c**) exclusively use small patches with high resource concentration. Note the 'microhabitat' separation in space of the two species' exclusive resources.

into subvolumes of size w^D . The volume w^D is the smallest food volume, or food-patch size P , consumed by the organism, that is, $w^D = P$. The total amount of food available to the organism is therefore the food contained in all subvolumes that it perceives as being filled with food, some of which are aggregated as larger food patches.

Larger species detect less total volume of food (only the larger patches) but can tolerate lower resource concentrations within their food, whereas smaller species detect more food (many small food patches), but require higher resource concentrations within it^{14,17–19}. If individuals of a species search k sub-volumes of size P in a time period dt , and resources are instantaneously replaced following consumption, the population growth rate of the species can be described as $dN/dt = qN(kRBP - L)$ (ref. 20), where N is population size, q converts resources into individuals and L is resource loss rate. Food is encountered within the habitat at a density $B = mw^F/w^D$. Resources within food are encountered at a concentration $R = rw^Q/w^F$. As w is proportional to organism size, L will reflect the balance between greater metabolic rate and lower mortality for larger organisms, and be approximately size-invariant²¹. A species can persist if $kRBP \geq L$, which requires that

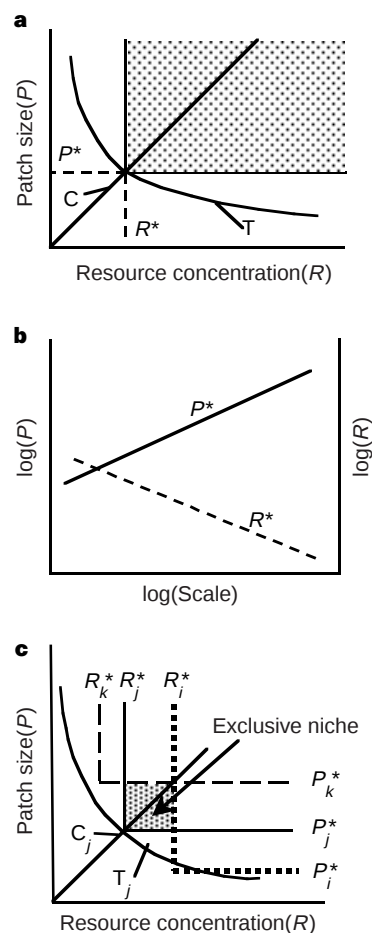


Figure 2 Graphical representation of the conditions for species persistence and coexistence. **a**, Minimum threshold patch size P^* and resource concentration R^* for a species. Increasing P implies that R can be reduced, yielding a trade-off relationship (curve T). However, larger patch sizes must be selected to encounter higher resource concentrations; that is, P is constrained to scale positively with R (line C). The intersection of curve T and line C yields the threshold P^* , R^* (equation (1)), which define suitable patches for the species (shaded area). **b**, Power law relationships for minimum food-patch size P^* and resource concentration R^* as a function of a species' size. **c**, Thresholds (R_i^* , P_i^*) for three species i, j, k that define the exclusive niche for species j (shaded area). Thresholds for species i and k will be positioned so that $P_j^* < P_k^*$ for every $R_j^* > R_i^*$ and to ensure that niche size $(P_k^* - P_j^*)(R_i^* - R_j^*)$ is large enough to ensure consumption rate $\geq$ loss rate.

a minimum food patch size (P^*) and minimum resource concentration (R^*) are exceeded. Greater P means that a lower R will meet resource losses, yielding a trade-off relationship: $P = Lw^{D-F}/(mkR)$ (curve T in Fig. 2a). However, larger organisms will encounter greater mean patch size but lower mean resource concentration in those patches. This encounter trade-off yields an 'encounter constraint' on the P and R that will satisfy resource losses. Substituting $R = rw^Q/w^F$ into the trade-off relationship yields the scaling law $P = Lw^{D-Q}/mk$ and substituting $P = w^D$ yields $R = Lw^{-F}/mk$. Recognizing that the scaling law for R is imbedded in the scaling law for P yields the encounter constraint $P = R(1/r)w^{F+D-Q}$ (line C in Fig. 2a). Unique thresholds P^* and R^* emerge from the intersection of the trade-off relationship and encounter constraint. These are simply the square root of, for P^* , the product of patch size for replacing losses and expected patch size encountered, and, for R^* , the product of resource concentration for replacing resource losses and expected resource concentration.

$$\begin{aligned} P^* &= (L/mkr)^{1/2} w^{D-Q/2} \\ R^* &= (Lr/mk)^{1/2} w^{Q/2-F} \end{aligned} \quad (1)$$

Because $D \geq F \geq Q$, P^* scales positively with size, whereas R^* scales negatively (Fig. 2b).

Applying these scaling laws to a group of species using similar resources, we find a 'packing' rule for how close in size species can be, that is, the size ratio γ between species of adjacent size. For a set of species i, j, k , ranked by increasing species size, the P^* and R^* of all three species define an exclusive niche for species j (Figs 1b, c, 2c): that is, both $P_i^* < P_j^* < P_k^*$ and $R_k^* < R_j^* < R_i^*$. For species j , there is a unique set of patches, ranging from size P_k^* to P_j^* and resource

concentration R_i^* to R_j^* , that are both too small for species k and too low in resource concentration for species i . If the species is to persist regardless of the abundance of competitor species, the resource intake rate in time dt from these exclusive food patches must equal resource losses: $k(P_k^* - P_i^*)(R_i^* - R_j^*) = L$. We can now find γ by assuming that the size ratios for each of the two adjacent species pairs are equal ($\gamma = w_k/w_j = w_j/w_i$), and replacing w_k with γw_j and w_i with $(1/\gamma)w_j$ in equation (1) so that $P_k^* = (L/mkr)^{1/2}(\gamma w_j)^{D-Q/2}$ and $R_i^* = (Lr/mk)^{1/2}(w_j/\gamma)^{Q/2-F}$. The ratio γ will probably deviate from 1 by less than an order of magnitude, so $\gamma^{D-Q/2} \approx \gamma^{F-Q/2}$. With this assumption, we can substitute functions of γ for P_k^* and R_i^* , and, using equation (1) for P_j^* and R_j^* , we can approximately solve for γ as a function of species' size:

$$\gamma(w) \approx \{1 + m^{1/2} w^{(F-D)/2}\}^{1/(D-Q/2)} \quad (2)$$

The body-size ratio $\gamma(w)$ should therefore decline with increasing organism size (Fig. 3a). This is true because $D \geq F \geq Q$, and because small resource-rich patches needed by smaller species occupy proportionately less total volume than larger, resource-poor patches used by larger species. To test this size-ratio prediction, we analysed body size patterns for two guilds of species that use similar resources: co-occurring, East African grazing mammals that all eat primarily herbaceous plants²²⁻²⁴ and vascular plants that compete for light in a Minnesota (USA) oak savanna²⁵. Instead of being constant^{1,5,11}, size ratios in these very different assemblages declined significantly with increasing size (Fig. 3b, c), and the relationships fit the predicted shape (equation (2); Fig. 3a).

The functional equation for size ratios (equation (2)) dictates the number of species ranging in size from $w_{\min}$ to $w_{\max}$ that can be 'packed' into an environment. The maximum size ($w_{\max}$) is determined by whether there is at least one suitable patch of size P^* and resource concentration R^* in a finite space of extent x . The number of suitable patches is found by dividing the total volume of resources rx^Q by the resource volume contained within suitable patches (P^*R^*). As P^* and R^* scale with w , however, the actual number of patches in the finite space is weighted by the probability of the occurrence of a patch with the length w specified by P^* and R^* . This probability is $(F-1)w^{-F}$ (ref. 19). Therefore, $(F-1)w^{-F}rx^Q/(P^*R^*) = 1$. Substituting for P^* and R^* (equation (1)) and solving for w yields

$$w_{\max} = [(F-1)x^Q krm/L]^{1/D} \quad (3)$$

A minimum resolution for a species within its environment ($w_{\min}$) may ultimately be set by physical constraints to a particular body plan or prey size (for example, vertebrates, plankton and so on).

Species richness (S) is then the number of exclusive niches allowed between $w_{\min}$ and $w_{\max}$ and is defined by

$$\prod_{i=1}^S \gamma(w_i) = \frac{w_{\max}}{w_{\min}} \quad (4)$$

which yields an approximate solution for S

$$S \approx \ln(w_{\max})/[\gamma(w_{\text{avg}})\ln(w_{\min})] \quad (5)$$

where w_{avg} is the mean body size in the guild. The functions implicit in $w_{\max}$ (equation (3)) and $\gamma(w_{\text{avg}})$ mean that the model also incorporates the effects on species richness of sampling area (x^2), habitat fragmentation (D) and the amount and distribution of food and resources (m, r, F, Q).

This model yields two unexpected predictions. First, it predicts a left-skewed, unimodal distribution of species richness versus organism size (Fig. 3d). This distribution reflects the larger size ratios and thus looser species packing required for smaller species (equation (2)) and the limitation of the largest species by the maximum patch size in the environment. The species richness-size distributions of both the East African herbivores and Minnesota plants are both significantly left-skewed (Fig. 3e, f), and differ from the log-normal or right-skewed distributions most commonly reported for species

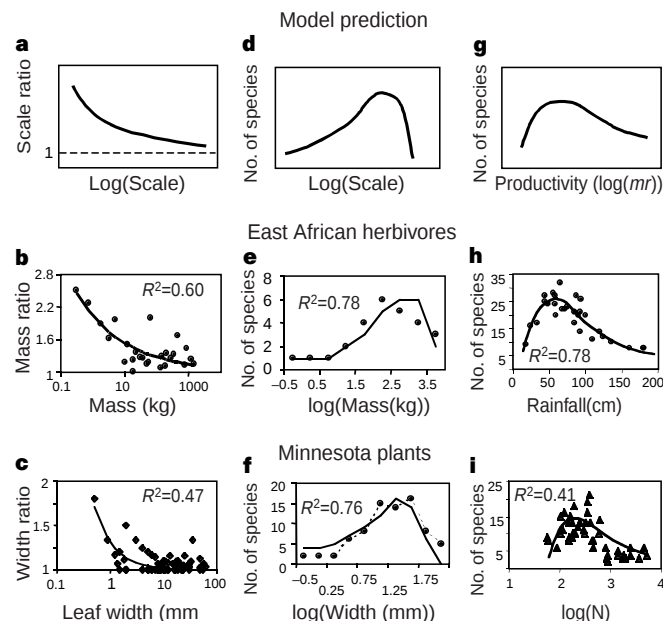


Figure 3 Predictions and tests of the scaling law model. **a**, Predicted size ratios (larger/smaller) for species of adjacent size versus size of the larger species; **d**, predicted number of species expected for communities versus log(size); and **g**, predicted number of species versus productivity, or log(mr). **b**, **c**, Observed size ratios (larger/smaller) of pairs of adjacent-sized species versus size of the larger of the pair. Size is body mass (kg) for Serengeti mammalian herbivores and leaf width (mm) for Minnesota plants. **e**, **f**, Observed frequency distributions of species richness versus classes of log(size) (dashed lines, closed circles). Both distributions are significantly left-skewed (D'Agostino test statistic, herbivores: -2.14 ; plants: -2.91 , $P < 0.05$). **h**, Observed number of herbivore species versus annual precipitation for 28 different preserves in East Africa; and **i**, number of plant species per 0.9 m^2 across a gradient of available soil NH_4^+ plus NO_3^- . Solid lines (with R^2) in **b**, **c**, **e**, **f**, **h** and **i** are least-squares fits to nonlinear functions predicted by scaling laws.

grouped by taxa or biogeographic region^{7,12,13}. Our model may not apply to communities that include species that use different resources or different habitats. Virtually all observed log-normal distributions combine diversity-size distributions of separate guilds (for example, nectarivores, granivores, herbivores, carnivores)¹³ or species from different habitats¹⁴.

Equation (5) also predicts the most commonly observed uni-modal pattern of species richness versus productivity^{8–10,26}, namely that species richness should increase rapidly and then decline gradually in response to increased productivity (represented as $\log(mr)$, Fig. 3g). As resources become more abundant, maximum patch size rapidly increases to allow larger species to exist. However, further increases in resource abundance cause food patches to coalesce, eliminating small, resource-rich patches and requiring greater size separation among smaller species. Once again, the model's predictions are supported for the mammalian herbivore and plant communities we examined (Fig. 3h, i).

The application of spatial scaling laws indicates that many of the mechanisms controlling biodiversity may emerge from first principles of how organisms find resources in space. The analysis formalizes earlier ideas that diversity depends on the number of spatial niches^{2,5,7,11}, and indicates that coexisting species cannot infinitely partition space^{9,10}. In addition, the model synthesizes recent ideas about how resource acquisition^{12,14,27} and spatial characteristics of habitat^{6,28} influence diversity. Clearly, other factors, including diversity of resource types²⁹, disturbance⁴, colonization limitation^{8,30} and biogeographical history^{3,10,30}, are also important in explaining diversity patterns. Nevertheless, the spatial scaling of resource use by species of different body size may explain many species-diversity patterns across a range of spatial scales and taxa. □

Methods

Size ratios (γ) and species richness (S) versus size relationships are for body mass of 27 mammalian grazer species >0.3 kg inhabiting open grasslands in the Serengeti National Park^{22–24}, and leaf width of 85 vascular plant species in a 1 ha of Minnesota oak savanna (leaves selected randomly on each of 10 plants of each species)²⁵. The selected African savanna grazing mammals partition different parts of largely the same species of plants^{22–24}, and the Minnesota savanna plants potentially compete for light²⁵. The number of >0.3 kg mammalian herbivore species found in 28 East African wildlife preserves was related to annual precipitation (a surrogate of productivity) at each preserve²³. The August diversity of Minnesota plants was measured from biomass clipped in August 1989–1997 out of three $10\text{ cm} \times 3\text{ m}$ strips within six replicate 81 m^2 plots that received either 0 or $26\text{ g m}^{-2}\text{ NH}_4\text{NO}_3$ per year since 1982 (ref. 25). Available NH_4^+ and NO_3^- was measured with an Alpkem autoanalyser following 0.01 M KCl extractions of $2 \times 15\text{ cm}$ deep soil cores²⁵. The relationships between γ and the size of the larger of the species pair (to avoid negative autocorrelation), S and different size classes, and S and productivity (either rainfall or $\log(\text{soil ammonium} + \text{nitrate})$) were fitted to nonlinear functions predicted by the scaling model. The distributions of number of species versus $\log(\text{size})$ were tested for skewness with D'Agostino tests.

Received 24 February; accepted 4 July 1999.

1. Brown, J. H. Two decades of homage to Santa Rosalia: toward a general theory of diversity. *Am. Zool.* **21**, 877–888 (1981).
2. May, R. M. How many species are there on earth? *Science* **241**, 1441–1449 (1988).
3. Schluter, D. & Ricklefs, R. (eds) *Species Diversity in Ecological Communities: Historical and Geographical Perspectives* (Univ. Chicago Press, Chicago, 1994).
4. Huston, M. A. *Biological Diversity: The Coexistence of Species on Changing Landscapes* (Cambridge Univ. Press, Cambridge, 1994).
5. Hutchinson, G. E. Homage to Santa Rosalia, or why are there so many kinds of animals? *Am. Nat.* **93**, 145–159 (1959).
6. MacArthur, R. H. Environmental factors affecting bird species diversity. *Am. Nat.* **98**, 387–397 (1964).
7. Morse, D., Stork, N. E. & Lawton, J. H. Fractal dimension of vegetation and the distribution of arthropod body lengths. *Nature* **314**, 731–732 (1985).
8. Tilman, D. & Pacala, S. W. in *Species Diversity in Ecological Communities: Historical and Geographical Perspectives* (eds Schluter, D. & Ricklefs, R.) 13–25 (Univ. Chicago Press, Chicago, 1994).
9. Rosenzweig, M. L. & Abramsky, Z. in *Species Diversity in Ecological Communities: Historical and Geographical Perspectives* (eds Schluter, D. & Ricklefs, R.) 52–65 (Univ. Chicago Press, Chicago, 1994).
10. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
11. Hutchinson, G. E. & MacArthur, R. H. A theoretical model of size distributions among species of animals. *Am. Nat.* **93**, 117–125 (1959).
12. Brown, J. H., Marquet, P. & Taper, M. Evolution of body size: consequences of an energetic definition

of fitness. *Am. Nat.* **142**, 573–584 (1993).

13. Siemann, E., Tilman, D. & Haarstad, J. Insect species diversity, abundance and body size relationships. *Nature* **380**, 704–706 (1996).
14. Belovsky, G. E. Optimal foraging and community structure: the allometry of herbivore food selection and competition. *Evol. Ecol.* **11**, 641–672 (1997).
15. Milne, B. T. Spatial aggregation and neutral models in fractal landscapes. *Am. Nat.* **139**, 32–57 (1992).
16. Milne, B. T. in *Wildlife and Landscape Ecology: Effects of Pattern and Scale* (ed. Bissonette, J.) 32–69 (Springer, New York, 1997).
17. Mandelbrot, B. B. *The Fractal Geometry of Nature*. (Freeman, New York, 1983).
18. O'Neill, R. V., Milne, B. T., Turner, M. G. & Gardner, R. H. Resource utilization scales and landscape pattern. *Land. Ecol.* **2**, 63–69 (1988).
19. Ritchie, M. E. Scale-dependent foraging and patch choice in fractal environments. *Evol. Ecol.* **12**, 309–330 (1998).
20. Schoener, T. W. Population growth regulated by intraspecific competition for energy or time: some simple representations. *Theor. Pop. Biol.* **6**, 265–307 (1973).
21. Damuth, J. Population density and body size in mammals. *Nature* **290**, 699–700 (1981).
22. Sinclair, A. R. E. & Norton-Griffiths, M. (eds) *Serengeti: Dynamics of an Ecosystem* (Univ. Chicago, Chicago, 1979).
23. Kingdon, J. *East African Mammals: an Atlas of Evolution in Africa* Vol. 1–7 (Academic, New York, 1979).
24. Prins, H. H. T. & Olff, H. in *Dynamics of Tropical Communities* (eds Newbery, D. M., Prins, H. H. T. & Brown, N. D.) 449–490 (Blackwell, Oxford, 1998).
25. Ritchie, M. E., Tilman, D. & Knops, J. M. H. Effects of herbivores on plant and nitrogen dynamics in oak savanna. *Ecology* **79**, 165–177 (1998).
26. Grime, J. P. *Plant Strategies and Vegetation Processes* (Wiley, Chichester, 1979).
27. Wright, D. H. Species-energy theory: an extension of species-area theory. *Oikos* **41**, 496–506 (1983).
28. Palmer, M. W. The coexistence of species in fractal landscapes. *Am. Nat.* **139**, 375–397 (1992).
29. Jones, C. G. & Lawton, J. H. Plant chemistry and insect species richness of British umbellifers. *J. Anim. Ecol.* **60**, 767–777 (1991).
30. Hubbell, S. P. A unified theory of biogeography and relative species abundance and its application to tropical rainforests and coral reefs. *Coral Reefs* **16**, S9–S21 (1997).

Acknowledgements. We thank B. Milne, J. H. Brown, H. DeKroon, J. H. Emlen, S. Grippe, A. Guss, N. Haddad, L. Li, S. Naeem and W. Pitt for comments. Supported by the US NSF, The Netherlands NWO, Santa Fe Institute and USU Ecology Center.

Correspondence and requests for information should be addressed to M.R. (e-mail: ritchie@cc.usu.edu).

A mesoscale approach to extinction risk in fragmented habitats

Renato Casagrandi & Marino Gatto

Dipartimento di Elettronica e Informazione, Politecnico di Milano, Via Ponzio, 34/5, 20133 Milano, Italy

Assessing the fate of species endangered by habitat fragmentation^{1–3} using spatially explicit and individual-based models^{4–7} can be cumbersome and requires detailed ecological information that is often unavailable. Conversely, Levins-like⁸ macroscale models^{9,10} neglect data on the distribution of local numbers, which are frequently collected by field ecologists^{11–13}. Here we present an alternative, mesoscale approach for metapopulations that are subject to demographic stochasticity, environmental catastrophes and habitat loss. Starting from a model that accounts for discrete individuals in each patch and assumes a birth–death stochastic process with global dispersal^{14,15}, we use a negative-binomial approximation¹⁶ to derive equations for the probability of patch occupancy and the mean and variance of abundance in each occupied patch¹⁷. A simple bifurcation analysis¹⁸ can be run to assess extinction risk. Comparison with both the original model and a spatially explicit model with local dispersal proves that our approximation is very satisfactory. We determine the sensitivity of metapopulation persistence to patch size, catastrophe frequency and habitat loss, and show that good dispersers are affected more by habitat destruction than by environmental disasters.

We consider an infinite network of equal patches with a uniform rain of propagules. We define $p_i(t)$ as the probability that, at time t , an integer number i of individuals is present in a patch; ν_i , μ_i and D_i as the birth, death and dispersal rates per capita in a patch containing i individuals; and m as the occurrence rate of catastrophes that wipe out a whole patch^{19,20}. Because of dispersal mortality and/or inability to colonize, only a fraction a of the average number of

dispersers per patch ($S_p = \sum_{i=1}^{\infty} D_i p_i(t)$) will end up in a patch. Habitat loss can reduce a , which may be thus multiplied by a factor u .

By using Poisson-like assumptions^{14,15,21}, we can stipulate an infinite number of equations for $p_i(t)$, $i = 0, 1, 2, \dots$ and test whether the p_i values converge to some distribution. With constant dispersal ($D_i = D$), no environmental catastrophes ($m = 0$), no habitat loss ($u = 1$) and local logistic dynamics, that is, $v_i - \mu_i = r(1 - i/K)$ with $v_i + \mu_i = \beta + \gamma i$, there is a unique stable probability distribution corresponding to either metapopulation persistence ($p_0 < 1$) or extinction ($p_0 = 1$). Figure 1a shows the boundary line in parameter space (D , r) between persistence and extinction. It was obtained by determining whether there are multiple probability distributions (stable or unstable) at equilibrium (see ref. 21 for another criterion). Persistence corresponds to intermediate D : too little dispersal prevents the rescue of patches that become extinct (see ref. 22 for extinction times) because of demographic stochasticity, whereas too much dispersal disproportionately increases the probability of ending up in an unsuitable habitat. The dark grey region corresponds to extinction as predicted by the deterministic model describing the overall balance of birth, death and dispersal for the metapopulation, namely $dM/dt = rM(1 - M/K) - D(1 - a)M$, where M is the abundance per patch.

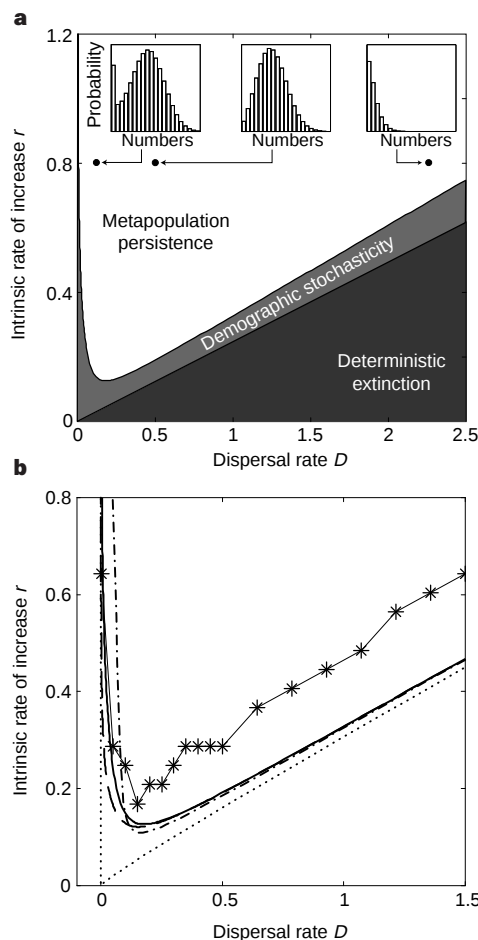


Figure 1 Boundaries separating persistence from extinction in parameter space (D , r) without catastrophes or habitat destruction ($m = 0$, $u = 1$). **a**, Functioning modes of model (1). The boxes show equilibrium distributions for p_i . Grey areas correspond to global extinction; dark grey $D > r/(1 - a)$, deterministic model; light grey, effect of demographic stochasticity. **b**, Mesoscale models versus the original model (solid line) and a spatially explicit model (asterisks). Mesoscale models are based on Poisson distribution (dotted line), negative binomial distribution with fixed (dash-dotted line) and varying (dashed line) clumping.

The set of infinite equations is not very manageable, because studying the conditions for persistence under different assumptions may be impractical. Deriving the persistence–extinction boundary with non-vanishing catastrophe frequency ($m > 0$) requires a computer-intensive numerical analysis, because there is no closed form for the probability distribution at equilibrium. We thus advocate the use of a sparing and ecologically meaningful approximation of the original model. This approximation is obtained by introducing the probabilities $\delta_i = p_i/(1 - p_0)$, conditional on a patch being occupied, so that the birth–death–dispersal–colonization process is given by

$$\frac{dp_0}{dt} = [-auS_\delta p_0 + (\mu_1 + D_1)\delta_1 + m](1 - p_0) \quad (1)$$

$$\begin{aligned} \frac{d\delta_i}{dt} = & [v_{i-1}(i-1) + au(1-p_0)S_\delta]\delta_{i-1} + \\ & -[(v_i + \mu_i + D_i)i + auS_\delta - (\mu_1 + D_1)\delta_1]\delta_i \\ & + (\mu_{i+1} + D_{i+1})(i+1)\delta_{i+1} \end{aligned} \quad (2)$$

where $S_\delta = S_p/(1 - p_0)$. Equation (1), describing the dynamics of empty patches, is Levins-like^{8–10} and incorporates the rescue effect and the extinction, due to death or emigration, of local populations consisting of one individual. From equation (2), describing the dynamics of numbers in occupied patches, we derive the dynamics of the mean and the variance of local abundance. Making specific assumptions about the density dependence of v_i , μ_i and D_i , we obtain equations for the moments of δ_i , in particular the conditional mean $M_\delta = \sum_{i=1}^{\infty} i\delta_i(t)$ and conditional variance σ_δ^2 . We reduce model (1) to two or three equations by assuming that, at any t , the δ_i values are approximated by a theoretical discrete distribution²³. Using distributions characterized by the first moment only provides δ_1 , σ_δ^2 and the skewness A_δ as functions of M_δ . Thus, two equations can specify the dynamics of local extinction frequency and of average abundance in non-empty patches (see Box 1 for examples). If we select a theoretical distribution characterized by the second moment as well, we obtain a three-dimensional model including the

Box 1 Two-dimensional mesoscale model describing the dynamics of patch occupancy and local mean abundance

Mesoscale metapopulation models

$$\begin{cases} \dot{p}_0 = (1 - p_0)[(\mu_1 + D)\delta_1 - auDp_0\Phi + m] \\ \dot{M}_\delta = M_\delta R(M_\delta) - \frac{r}{K}\Psi + (\mu_1 + D)\delta_1 M_\delta + \\ - [D(1 - au) + auDp_0M_\delta]\Phi \end{cases}$$

Constant dispersal ($D_i = D$)

Logistic Allee effect

$$\begin{aligned} \Phi &= M_\delta & \Phi &= M_\delta \\ R &= r\left(1 - \frac{M_\delta}{K}\right) & R &= r\left(1 - \frac{M_\delta}{K}\right)\left(\frac{M_\delta}{\epsilon} - 1\right) \\ \Psi &= \sigma_\delta^2 & \Psi &= \frac{A_\delta}{\epsilon} + 3\frac{M_\delta}{\epsilon}\sigma_\delta^2 - \left(1 + \frac{K}{\epsilon}\right)\sigma_\delta^2 \end{aligned}$$

Density dependent dispersal ($D_i = D \cdot i$)

$$\begin{aligned} \Phi &= M_\delta^2 + \sigma_\delta^2 & \Phi &= M_\delta^2 + \sigma_\delta^2 \\ R &= r\left(1 - \frac{M_\delta}{K}\right) & R &= r\left(1 - \frac{M_\delta}{K}\right)\left(\frac{M_\delta}{\epsilon} - 1\right) \\ \Psi &= \sigma_\delta^2 & \Psi &= \frac{A_\delta}{\epsilon} + 3\frac{M_\delta}{\epsilon}\sigma_\delta^2 - \left(1 + \frac{K}{\epsilon}\right)\sigma_\delta^2 \end{aligned}$$

As an example, we show analytical expressions for different types of density dependence and dispersal. See text for definition of δ_i , σ_δ^2 and A_δ . They depend on average abundance according to different theoretical distributions (see Methods).

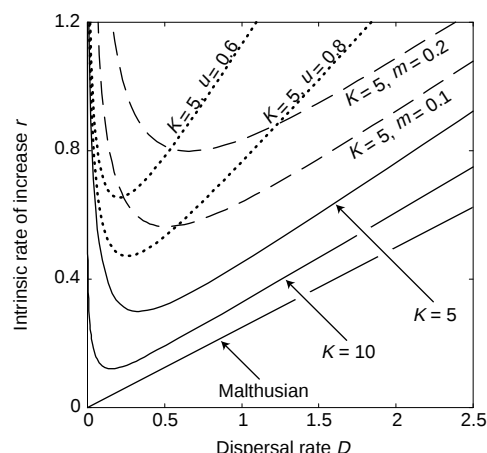


Figure 2 The influence of key ecological parameters on persistence–extinction boundaries (calculated by bifurcation analysis) for three-dimensional mesoscale models (negative binomial with varying clumping). Shown are the effect of decreasing local carrying capacities K (solid line); the further effect of different occurrence rates m of environmental disasters (dashed line) and levels of habitat loss (dotted line, u = fraction of undestroyed patches).

dynamics of abundance variance. Figure 1b compares the performance of the original and the approximate models in terms of extinction–persistence boundaries. Those of reduced models follow from a simple stability analysis. Although the Poisson approximation is too crude, the negative binomial with fixed (varying) clumping can satisfactorily (well) mimic the properties of the original model. Fixed clumping might be estimated from distributional data^{11,12}.

Explorations with stochastic cellular automata²⁴, considering a finite, two-dimensional array of patches and local dispersal, show that the basic features of stability boundaries obtained by the negative-binomial approximation are unchanged (Fig. 1b). Thus, of the two properties ‘being discrete’ and ‘being spatial’²⁵, the first is indeed crucial if we aim to assess metapopulation persistence in a simple, effective and flexible way. With the mesoscale models we can greatly reduce computation times and readily establish the sensitivity of results to different metapopulation parameters, because bifurcation analysis¹⁸ of three ODEs requires a few seconds. For instance, Fig. 2 shows the impact on extinction risk of patch size K , catastrophe frequency m and percentage of preserved habitat u .

We summarize the contribution of the different forces driving extinction in Fig. 3. If, besides demographic stochasticity, environmental catastrophes occur at a substantial rate, the lower dispersal

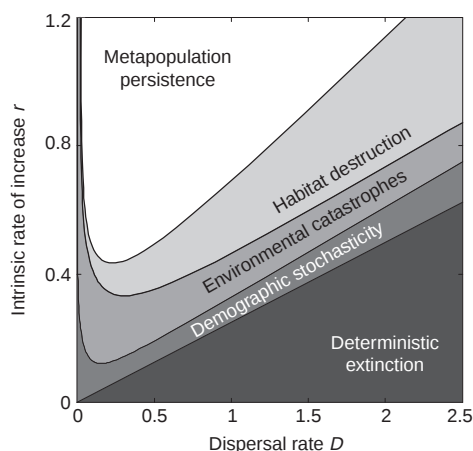


Figure 3 Summary of various contributions to metapopulation extinction in parameter space (D , r). Carrying capacity $K = 10$, environmental disaster rate $m = 0.1$ and fraction of undestroyed habitat $u = 0.8$.

threshold necessary for persistence increases conspicuously, and the upper threshold decreases. If man destroys suitable habitat, this reduces the upper dispersal threshold. Organisms that would survive in a pristine but patchy habitat by their evolved ability to disperse will become extinct in a partly destroyed landscape because they are lost to unsuitable habitat by that very ability. In fact, studies on bird conservation in Europe²⁶ show that modification of the agricultural landscape is one of the most relevant causes of declining populations. □

Methods

We obtain the boundary lines shown in Figs 1, 2 and 3, where not specified otherwise, by setting: $a = 0.75$, $K = 10$, $\beta = 2$, $\gamma = 0.001$, $m = 0$ and $u = 1$. For the Poisson distribution, $\sigma_s^2 = M_s$ and $\delta_1 = e^{-(M_s - 1)}$; for the negative binomial distribution, $\sigma_s^2 = (M_s - 1) + ((M_s - 1)^2/h)$ and $\delta_1 = (1 + (M_s - 1/h))^{-h}$, where the clumping h is either fixed ($h = 0.7$) or varying according to the relationship linking the third moment with the first two moments²⁷. The spatially explicit model is a 6×6 cellular automaton with local dispersal (four neighbours on a torus). For each patch, we draw dispersal, birth and death at random. Birth and death probabilities are logistically dependent upon local abundance. For each point of a grid in parameter space, we run simulations several times over a time horizon equal to $100/r$. The stopping rule for the simulations (metapopulation is considered either persistent or prone to extinction and no further simulation is run) is given by a sequential statistical test²⁸ (100% persistence probability is tested against 80% persistence probability, type I error = 5%, type II error = 10%).

Received 31 March; accepted 3 June 1999.

- Bascompte, J. & Solé, R. V. Habitat fragmentation and extinction thresholds in spatially explicit models. *J. Anim. Ecol.* **65**, 465–473 (1996).
- Hanski, I., Gilpin, M. (eds) *Metapopulation Biology: Ecology, Genetics, and Evolution* (Academic, San Diego, California, 1997).
- Hanski, I. Metapopulation dynamics. *Nature* **396**, 41–50 (1998).
- Tilman, D. & Kareiva, P. *Spatial Ecology* (Princeton Univ. Press, New Jersey, 1997).
- Bascompte, J. & Solé, R. V. *Modeling Spatiotemporal Dynamics in Ecology* (Springer, Berlin, 1998).
- De Angelis, D. L. & Gross, L. J. (eds) *Individual-Based Models and Approaches in Ecology* (Chapman & Hall, New York, 1992).
- Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature* **359**, 826–829 (1992).
- Levins, R. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull. Entomol. Soc. Am.* **15**, 237–240 (1969).
- Gotelli, N. J. Metapopulation models: The rescue effect, the propagule rain, and the core-satellite hypothesis. *Am. Nat.* **138**, 768–776 (1991).
- Day, J. R. & Possingham, H. P. A stochastic metapopulation model with variability in patch size and position. *Theor. Popul. Biol.* **48**, 333–360 (1995).
- den Boer, P. J. Density limits and survival of local populations in 64 carabid species with different power of dispersal. *J. Evol. Biol.* **3**, 19–48 (1990).
- Whitlock, M. C. Nonequilibrium population structure in forked fungus beetles: extinction, colonization and the genetic variance among populations. *Am. Nat.* **139**, 952–970 (1992).
- Crawley, M. J. & Brown, S. L. Seed limitation and the dynamics of feral oilseed rape on the M25 motorway. *Proc. R. Soc. Lond. B* **259**, 49–54 (1995).
- Chesson, P. L. Models for spatially distributed populations: The effect of within-patch variability. *Theor. Popul. Biol.* **19**, 288–325 (1981).
- Drechsler, M. & Wissel, C. Separability of local and regional dynamics in metapopulations. *Theor. Popul. Biol.* **51**, 9–21 (1997).
- Anderson, R. M. & May, R. M. Regulation and stability of host–parasite population interactions—I. Regulatory processes. *J. Anim. Ecol.* **47**, 219–247 (1978).
- Bolker, B. M. & Pacala, S. W. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Popul. Biol.* **52**, 179–197 (1997).
- Kuznetsov, Y. A. *Elements of Applied Bifurcation Theory* (Springer, New York, 1995).
- Kareiva, P. & Wennergren, V. Connecting landscape patterns to ecosystem and population process. *Nature* **373**, 299–302 (1995).
- Olivieri, I., Michalakis, Y. & Gouyon, P. H. Metapopulation genetics and the evolution of dispersal. *Am. Nat.* **146**, 202–228 (1995).
- Chesson, P. L. Persistence of a Markovian population in a patchy environment. *Z. Wahrscheinlichkeitstheorie* **66**, 97–107 (1984).
- Lande, R. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *Am. Nat.* **142**, 911–927 (1993).
- Chesson, P. L. in *Modeling Spatiotemporal Dynamics in Ecology* (eds Bascompte, J. & Solé, R. V.) (Springer, Berlin, 1998).
- Wolfram, S. *Cellular Automata and Complexity: Collected Papers* (Addison-Wesley, Reading, Massachusetts, 1994).
- Durrett, R. & Levin, S. The importance of being discrete (and spatial). *Theor. Popul. Biol.* **46**, 363–364 (1994).
- Tucker, G. M. & Heath, M. F. *Birds in Europe. Their Conservation Status* (Birdlife International, Cambridge, 1994).
- Pugliese, A., Rosà, R. & Damaggio, M. L. Analysis of a model for macroparasitic infection with variable aggregation and clumped infections. *J. Math. Biol.* **36**, 419–447 (1998).
- Kendall, M. & Stuart, A. *The Advanced Theory of Statistics* (Griffin, London, 1977).

Acknowledgements. We thank G. De Leo, C. Dimou and I. Hanski for suggestions. We thank Politecnico di Milano for support.

Correspondence and requests for materials should be addressed to M.G. (e-mail: gatto@elet.polimi.it).

Mesopredator release and avifaunal extinctions in a fragmented system

Kevin R. Crooks* & Michael E. Soulé†

* Department of Biology, University of California, Santa Cruz, California 95064, USA

† The Wildlands Project, PO Box 1302 2010, Hotchkiss, Colorado 81419, USA

Mammalian carnivores are particularly vulnerable to extinction in fragmented landscapes¹, and their disappearance may lead to increased numbers of smaller carnivores that are principle predators of birds and other small vertebrates. Such 'mesopredator release'² has been implicated in the decline and extinction of prey species²⁻⁶. Because experimental manipulation of carnivores is logistically, financially and ethically problematic^{6,7}, however, few studies have evaluated how trophic cascades generated by the decline of dominant predators combine with other fragmentation effects to influence species diversity in terrestrial systems. Although the mesopredator release hypothesis has received only limited critical evaluation⁸ and remains controversial⁹, it has become the basis for conservation programmes justifying the protection of carnivores⁶. Here we describe a study that exploits spatial and temporal variation in the distribution and abundance of an apex predator, the coyote, in a landscape fragmented by development. It appears that the decline and disappearance of the coyote, in conjunction with the effects of habitat fragmentation, affect the distribution and abundance of smaller carnivores and the persistence of their avian prey.

In coastal southern California, intensive urbanization over the past century has destroyed most of the native sage-scrub habitat, leaving undeveloped steep-sided canyons as habitat islands in an urban sea. The mesopredator release hypothesis² was proposed as a possible mechanism to explain the rapid disappearance of scrub-breeding birds from this system. It predicted that the decline of the most common large predator (coyote) would result in the ecological release of native (striped skunk, raccoon, grey fox) and exotic (domestic cat, opossum) mesopredators, and that increased predation by these effective predators^{5,10-12} would result in higher mortality and local extinction rates of scrub-breeding birds.

To test these predictions, we surveyed coyotes, mesopredators and scrub-breeding birds in 28 urban habitat fragments (see Methods). Coyote populations have declined or disappeared from some fragments; backward elimination multiple regression (BEMR) analyses (Table 1a) indicated that fragment size was a positive predictor of mean coyote abundance (averaged over quarterly sampling sessions). As predicted, the relationship between coyote and mesopredator abundance among fragments was consistently negative (Table 2). Total mesopredator abundance, summed over all mesopredator species, was higher in fragments with fewer coyotes; coyote abundance had the strongest negative relationship with grey fox, cat and opossum abundance (Table 2). BEMR analyses indicated that coyote abundance was the strongest predictor of total mesopredator, fox and opossum abundance after accounting for the potentially confounding effects of fragment area, age and isolation (Table 1b). The most important predictor of cat abundance was the inverse of fragment area, as would be expected because smaller fragments have proportionally more urban edge and therefore greater access by housecats bordering the fragment.

Simply the presence or absence of coyotes in a fragment also influenced mesopredator abundance. Mean total mesopredator abundance was more than twice as high in fragments that coyotes never visited during the course of the study (mean, 1.17; s.d., 0.299)

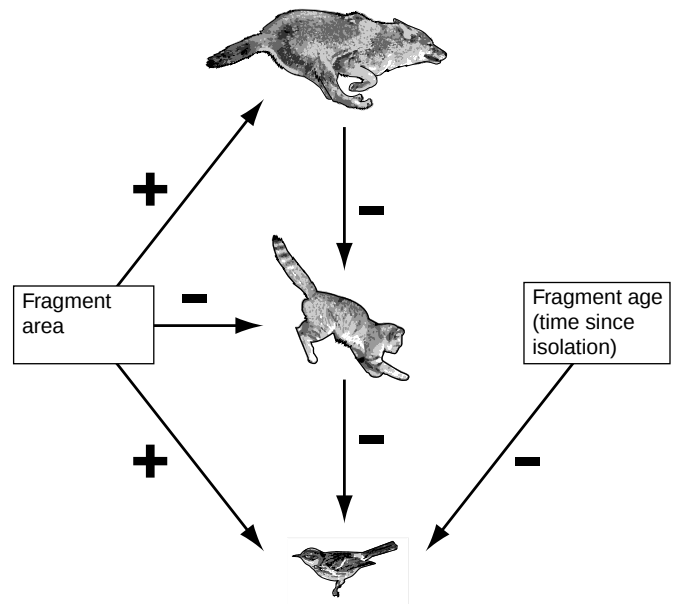


Figure 1 Model of the combined effects of trophic cascades and island biogeographical processes on top predators (for example, coyote), mesopredators (domestic cat) and prey (scrub-breeding birds) in a fragmented system. Direction of the interaction is indicated with a plus or minus.

than in fragments where coyotes were detected at least once (mean, 0.52; s.d., 0.436). Coyote presence had the strongest negative effect on domestic cat, opossum and raccoon abundance (Table 2). BEMR indicated that coyote presence or absence was an important predictor of total mesopredator, domestic cat, opossum and raccoon abundance after accounting for fragment area, age and isolation (Table 1c); the negative effect of fragment area was retained in the final regression models for total mesopredator, grey fox and domestic cat abundance.

In accordance with the mesopredator release hypothesis, the direction of the correlation between the number of native scrub-specialist bird species persisting in fragments (see Methods) and mesopredator abundance was consistently negative (Table 2). Bird species diversity decreased with total mesopredator abundance; bird diversity had the strongest inverse correlation with grey fox, domestic cat, opossum and raccoon abundance (Table 2). BEMR revealed that the positive effect of fragment area and the negative effect of fragment age were the strongest determinants of bird diversity in this system (Table 1d). However, the negative effects of total mesopredator, cat and raccoon abundance on bird diversity persisted even after accounting for age and area effects.

According to the mesopredator release hypothesis, the top predator should have an indirect and positive effect on bird species diversity¹³. As predicted, scrub bird diversity was higher in fragments where coyotes were either present or more abundant (Table 2). Both coyote presence or absence and coyote abundance remained significant predictors of bird diversity even after accounting for the strong effects of fragment area and age on bird populations (Table 1d).

Coyote abundance also varied across time in some fragments, permitting a more direct test of the causal mechanisms underlying the correlational patterns observed in the regression analyses; temporal variability was masked in the above analyses by using mean abundances averaged across quarterly sampling sessions in each fragment. In the 13 fragments that coyotes visited only temporarily during the study, mean abundance of total mesopredators in quarters without coyotes was higher than in quarters with coyotes (Wilcoxon matched pairs test: $Z = 3.110$, $P = 0.002$); this pattern of temporal avoidance was significant for foxes ($Z = 2.667$, $P = 0.008$), cats ($Z = 2.353$, $P = 0.019$) and skunks ($Z = 2.045$,

$P = 0.041$), and was not significant for opossums ($Z = 1.334$, $P = 0.182$) or raccoons ($Z = 1.007$, $P = 0.314$). Indeed, temporal variance in total mesopredator visitation rate was significantly higher in the 13 fragments in which coyotes came and went compared to the 15 fragments in which coyotes were either constantly present or absent ($t = 2.18$, $P = 0.038$). Finally, within each of the five fragments surveyed for two years, total mesopredator visitation rate increased when coyote visitations declined (Table 3); this temporal avoidance between coyote and mesopredators was largely driven by coyote–cat interactions.

Mesopredators not only temporally avoided coyotes within fragments, but also avoided sites in fragments where coyotes were most active. In the 11 fragments where mesopredators were detected and

where coyotes were present in every quarterly sampling session, coyotes and mesopredators visited the same track station on the same night significantly less than expected based on random visitations of both taxa (contingency $X^2 = 12.39$, $P < 0.001$). This pattern was evident for foxes ($X^2 = 4.572$, $P = 0.032$ in 8 fragments with foxes) and opossums ($X^2 = 2.96$, $P = 0.086$ in 9 fragments), but was not significant for cats ($X^2 = 0.856$, $P = 0.355$ in 11 fragments), skunks ($X^2 = 1.74$, $P = 0.187$ in 7 fragments) or raccoons ($X^2 = 0.900$, $P = 0.343$ in 4 fragments).

The interactions between coyotes, cats and birds probably have the strongest impact on the decline and extinction of scrub-breeding birds. Coyotes kill domestic cats in these habitat fragments. Cat remains were found in most fragments with coyotes, and 21% of 219

Table 1 BEMR models of effects of trophic interactions and biogeographical variables

	R^2	Whole model P	Parameter estimate	P
(a) Dependent variable: coyote abundance*				
Coyote abundance	0.111	0.046		
Area			0.381	0.046
(b) Dependent variables: mesopredator abundance†				
Total mesopredator abundance	0.314	0.002		
Coyote abundance			−0.560	0.002
Grey fox abundance	0.356	<0.001		
Coyote abundance			−0.596	<0.001
Domestic cat abundance	0.316	0.002		
Area			−0.562	0.002
Opossum abundance	0.373	<0.001		
Coyote abundance			−0.611	<0.001
Skunk abundance				
n.s.‡				
Raccoon abundance				
n.s.				
(c) Dependent variables: mesopredator abundance§				
Total mesopredator abundance	0.266	0.021		
Coyote presence/absence			−0.349	0.063
Area			−0.290	0.119
Grey fox abundance	0.083	0.137		
Area			−0.288	0.137
Domestic cat abundance	0.475	<0.001		
Coyote presence/absence			−0.418	0.011
Area			−0.438	0.008
Opossum abundance	0.285	0.003		
Coyote presence/absence			−0.533	0.003
Skunk abundance				
n.s.				
Raccoon abundance	0.122	0.068		
Coyote presence/absence			−0.350	0.068
(d) Dependent variable: bird diversity 				
Bird diversity	0.886	<0.001		
Coyote presence/absence			0.464	<0.001
Area			0.511	<0.001
Age			−0.388	<0.001
Bird diversity	0.757	<0.001		
Coyote abundance			0.234	0.042
Area			0.558	<0.001
Age			−0.531	<0.001
Bird diversity	0.741	<0.001		
Total mesopredator abundance			−0.199	0.101
Area			0.569	<0.001
Age			−0.486	<0.001
Bird diversity	0.745	<0.001		
Domestic cat abundance			−0.235	0.082
Area			0.516	<0.001
Age			−0.480	<0.001
Bird diversity	0.762	<0.001		
Raccoon abundance			−0.241	0.031
Area			0.592	<0.001
Age			−0.478	<0.001
Bird diversity¶	0.739	<0.001		
Area			0.648	<0.001
Age			−0.558	<0.001

At successive steps in the backward elimination procedure, the least significant independent variable was removed from the model if the significance of the parameter estimate >0.15 (refs 18, 19). This process was continued until no variables met this criteria. Tolerance values indicated that no set of independent variables violated multicollinearity assumptions¹⁹. The resulting final models are presented above, with the independent variables retained in the final model in italics. Path analyses conducted on these data yielded similar results as the multiple regression analyses.

* Independent variables: fragment area, fragment age and fragment isolation.

† Independent variables: coyote abundance, fragment area, fragment age and fragment isolation.

§ Independent variables: coyote presence/absence, fragment area, fragment age and fragment isolation.

|| Independent variables: coyote presence/absence, coyote abundance or mesopredator abundance, and fragment area, age and isolation.

‡ No independent variables were retained in the final model.

¶ Skunk, fox and opossum abundance were not retained in final models that included area and age effects.

coyote scats collected in these sites contained cat remains. Moreover, 25% of radio-collared cats were killed by coyotes (K.C., manuscript in preparation). Perhaps the strongest effect of coyotes on cats, however, is indirect. Seventy-one per cent of 636 respondents to questionnaires distributed to residents bordering the fragments realized that coyotes were a threat to cats, 42% of all cat owners in areas with coyotes reported that coyotes had attacked or killed their cats and, most importantly, 46% of cat owners restricted their cat's outdoor activity when they believed coyotes were in the fragment.

Unlike wild predators, domestic cats are recreational hunters maintained far above carrying capacity by nutritional subsidies from their owners; they continue to kill prey species even when populations of that species are low¹¹. Thirty-two per cent of residents bordering the San Diego fragments owned cats, and on average each cat owner owned 1.7 cats. Seventy-seven per cent of cat owners let their cats outdoors, and 84% of outdoor cats brought back kills to the residence. Thus, approximately 35 hunting, outdoor cats surround a moderately sized fragment (~20 ha) bordered by 100 residences. In comparison, each fragment may support only one or two pairs of native predators such as foxes or coyotes. Cat owners reported that each outdoor cat that hunted returned on average 24 rodents, 15 birds and 17 lizards to the residence each year. Using these data, we estimate that cats surrounding a moderately sized fragment (~100 residences) return about 840 rodents, 525 birds and 595 lizards to residences per year. These approximations are probably underestimates, assuming that cats do not bring back all prey that they kill¹⁴. Identification of 68 prey items returned by cats bordering the fragments indicated that 67% of 26 rodents, 95% of 21 birds and 100% of 11 lizards were native species.

This level of bird predation appears to be unsustainable. Existing population sizes of some birds do not exceed 10 individuals in small to moderately sized fragments¹⁵, so even modest increases in predation pressure from mesopredators, in conjunction with other fragmentation effects, may quickly drive native prey species, especially rare ones, to extinction. Extinctions of scrub-breeding birds are frequent and rapid; at least 75 local extinctions may have occurred in these fragments over the past century¹⁵.

Our results indicate that the disappearance of a dominant

carnivore results in elevated numbers and activity of mesopredators that exert strong predation pressure on native prey species. This conclusion is strengthened by changes in mesopredator activity in accord with temporal changes in coyote presence within fragments, as well as direct evidence of coyote predation on mesopredators and mesopredator predation on birds. We conclude that these trophic interactions combine with fragmentation effects to help structure this ecological community (Fig. 1). □

Methods

Biogeographical variables. We used fragment area, age and isolation as island biogeographical descriptors of the 28 urban habitat fragments². The total area of each fragment was taken from digitized images of scaled aerial photographs taken in 1995 (range: 2–102 ha). Fragment age was defined as the number of years since isolation of the fragment by urban development (range: 11–95 yr). Fragment isolation was measured as the distance to the closest fragment of equal or larger size (range: 40–2,865 m). Biogeographical variables were log-transformed for analyses.

Carnivore surveys. From September 1995 through to August 1997, we conducted carnivore surveys in 28 habitat fragments originally studied in ref. 2. Relative abundance for each species was determined by establishing track detection stations at 250-m intervals along transects in each fragment, and conducting track surveys for five consecutive days in the autumn, winter, spring and summer for one year. In five fragments where coyote presence varied during the first year of surveys, we extended surveys for a second year to monitor further the effects of variation in abundance within sites. The presence of each species was verified using scat and remotely triggered camera surveys. Abundance in each quarter was expressed as the total number of visits to track stations for each species divided by the total sampling effort^{16,17}; track indices were log-transformed for analyses. For each species, we averaged track indices across quarterly sampling sessions to derive a mean abundance per fragment for the duration of the study. In addition to calculating abundance for each species individually, we summed the relative abundance of mesopredators in each fragment to derive one metric for the total abundance of all small carnivores.

Bird surveys. We determined the number of scrub bird species in each fragment by point count and transect surveys, conducted in each fragment at least three different times by at least two different teams of trained observers from April 4 to June 9 1997 between sunrise and 10:30. Eight-minute point counts were conducted at stations established in or near native habitat at ~250-m intervals along the long axis of each fragment. For transect surveys, we walked slowly along the entire fragment and recorded all species detected (mean time spent per transect survey in each fragment, 107 min). We then combined the species occurrences generated by both the point count and transect surveys to calculate the number of scrub bird species at each site. We considered only those species that specialize on chaparral and coastal sage scrub habitat and rarely breed in developed sites: California quail, wren, spotted towhee, Bewick's wren, California thrasher, greater roadrunner, cactus wren and California gnatcatcher. Bird diversity was square-root transformed for analyses.

Received 22 February; accepted 5 July 1999.

1. Woodroffe, R. & Ginsberg, J. R. Edge effects and extinction of populations inside protected areas. *Science* **280**, 2126–2128 (1998).
2. Soulé, M. E. *et al.* Reconstructed dynamics of rapid extinctions of chaparral-requiring birds in urban habitat islands. *Conserv. Biol.* **2**, 75–92 (1988).
3. Sovada, M. A., Sargeant, A. B. & Grier, J. W. Differential effects of coyotes and red foxes on duck nest success. *J. Wildl. Mgmt* **59**, 1–9 (1995).

Table 2 Trophic interactions

	Coyote abundance <i>r</i> †	Coyote presence/ absence <i>t</i> ‡	Bird diversity <i>r</i> §
Total mesopredator abundance	–0.569***	–2.463**	–0.539***
Fox abundance	–0.597***	–1.090	–0.361*
Domestic cat abundance	–0.375**	–3.344***	–0.635***
Opossum abundance	–0.611***	–3.220***	–0.464**
Skunk abundance	–0.105	0.362	–0.112
Raccoon abundance	–0.264	–1.908*	–0.487***
Bird diversity	0.452**	5.580***	

† Pearson correlations between mean coyote abundance (averaged across quarterly sampling sessions) and mean abundance of mesopredator species (averaged across quarterly sampling sessions) or number of scrub-breeding bird species in each of the 28 habitat fragments.

‡ *t*-test of mean abundance of mesopredator species or bird species diversity as function of coyote presence or absence in each fragment.

§ Pearson correlations between number of scrub-breeding bird species per fragment and mean mesopredator abundance.

P* < 0.10, *P* < 0.05, ****P* < 0.01.

Table 3 Temporal avoidance of coyotes by mesopredators

Coyote abundance versus:	Total mesopredator abundance	Domestic cat abundance	Fox abundance	Skunk abundance	Opossum abundance	Raccoon abundance
Fragment	<i>r</i> †	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>
Baja	–0.785**	–0.667*	–0.577**	–4.401	–0.502	–0.328
Laurel	–0.769**	–0.814**	–0.426	–0.694*	–0.521	0.338
Spruce	–0.740**	–0.846***	–0.559	–0.272	–0.582	–0.198
Washington	–0.679*	–0.709**	–0.639	–0.424	–0.122	–0.773**
Titus	–0.723*	–0.424	–0.695*	–0.253	–0.581	0.193

† Pearson correlations between quarterly mesopredator and coyote abundance within five habitat fragments surveyed over two years.

P* < 0.10, *P* < 0.05, ****P* < 0.01.

- Palomares, F., Gaona, P., Ferreras, P. & Delibes, M. Positive effects on game species of top predators by controlling smaller predator populations: an example with lynx, mongooses, and rabbits. *Conserv. Biol.* **9**, 295–305 (1995).
- Rogers, C. M. & Caro, M. J. Song sparrows, top carnivores, and nest predation: a test of the mesopredator release hypothesis. *Oecologia* **116**, 227–233 (1998).
- Soulé, M. E. & Terborgh, J. *Continental Conservation: Scientific Foundations for Regional Reserve Networks* (Island, Washington, 1999).
- Estes, J. A. in *Linking Species and Ecosystems* (eds Jones, C. G. & Lawton, L. H.) 151–158 (Chapman and Hall, New York, 1995).
- Litvaitis, J. A. & Villafuerte, R. Intraguild predation, mesopredator release, and prey stability. *Conserv. Biol.* **10**, 676–677 (1996).
- Wright, S. J., Gompper, M. E. & DeLeon, B. Are large predators keystone species in neotropical forests—the evidence from Barro Colorado Island. *Oikos* **71**, 279–294 (1994).
- Wilcove, D. S. Nest predation in forest tracts and the decline of migratory songbirds. *Ecology* **66**, 1211–1214 (1985).
- Churcher, J. B. & Lawton, J. H. Predation by domestic cats in an English village. *J. Zool. (Lond.)* **212**, 439–456 (1987).
- Leimgruber, P., McShea, W. J. & Rappole, J. H. Predation on artificial nests in large forest blocks. *J. Wildl. Mgmt* **58**, 254–260 (1994).
- Fretwell, S. D. Food chain dynamics: the central theory of ecology? *Oikos* **50**, 291–301 (1987).
- George, W. Domestic cats as predators and factors in winter shortages of raptor prey. *Wilson Bull.* **86**, 384–396 (1974).
- Bolger, D. T., Alberts, A. & Soulé, M. E. Occurrence patterns of bird species in habitat fragments: sampling, extinction, and nested species subsets. *Am. Nat.* **137**, 155–166 (1991).
- Linhardt, S. B. & Knowlton, F. F. Determining the relative abundance of coyotes by scent station lines. *Wildl. Soc. Bull.* **3**, 119–124 (1975).
- Conner, M. C., Labisky, R. F. & Progulski, D. R. Jr Scent-station indices as measures of population abundance for bobcats, raccoons, gray foxes, and opossums. *Wildl. Soc. Bull.* **11**, 146–152 (1983).
- Bendel, R. B. & Afifi, A. A. Comparison of stepping rules in forward regression. *J. Am. Stat. Assoc.* **72**, 46–53 (1977).
- Tabachnick, B. G. & Fidell, L. S. *Using Multivariate Statistics* 3rd edn (HarperCollins College Publishers, New York, 1996).

Acknowledgements. We thank L. Angeloni, D. Bolger, T. Case, J. Crooks, D. Doak, J. Estes, R. Fisher, S. Hathaway, D. Menendez, S. Minta, P. Raimondi, B. Rice, and A. Suarez for their valuable help with this research, and C. Bell for illustrating Fig. 1. This work was funded by D. Brimm, an NSF Graduate Research Fellowship, an EPA STAR Fellowship and an American Society of Mammalogist grant (K.R.C.).

Correspondence and requests for materials should be addressed to K.R.C. (e-mail: krcrooks@earthlink.net.).

Linking a genetic defect to its cellular phenotype in a cardiac arrhythmia

Colleen E. Clancy* & Yoram Rudy*†

Cardiac Bioelectricity Research and Training Center, † Department of Biomedical Engineering and * Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, Ohio 44106-7207, USA

Advances in genetics and molecular biology have provided an extensive body of information on the structure and function of the elementary building blocks of living systems. Genetic defects in membrane ion channels can disrupt the delicate balance of dynamic interactions between the ion channels and the cellular environment, leading to altered cell function^{1–3}. As ion-channel defects are typically studied in isolated expression systems, away from the cellular environment where they function physiologically, a connection between molecular findings and the physiology and pathophysiology of the cell is rarely established. Here we describe a single-channel-based Markovian modelling approach that bridges this gap. We achieve this by determining the cellular arrhythmogenic consequences of a mutation in the cardiac sodium channel that can lead to a clinical arrhythmogenic disorder (the long-QT syndrome) and sudden cardiac death.

Several distinct genetic mutations in the *SCN5A* gene give rise to a congenital form of the long-QT syndrome and have been mapped to the α -subunit of the cardiac sodium channel (LQT3)⁴. The most severe is the Δ KPQ mutation, a three-amino-acid deletion of Lys 1505, Pro 1506 and Gln 1507 in the highly conserved portion of the III–IV linker, which is responsible for fast inactivation⁵. Clinically, the Δ KPQ mutation is associated with substantial prolongation of the Q–T interval on the electrocardiogram, which may precede syncope and sudden cardiac death.

To evaluate the electrophysiological consequences of the Δ KPQ defect at the level of the cardiac action potential, we constructed Markov models of the wild-type and Δ KPQ mutant channels based on experimental data^{5–7}. The models were then integrated into the Luo–Rudy theoretical model of the cardiac ventricular action potential^{8–10}.

The Markovian models for the wild-type and Δ KPQ sodium channel are shown in Fig. 1. The wild-type channel model (Fig. 1a) includes three closed states (C3, C2 and C1), a conducting open state (O), and fast and slow inactivation states (IF and IS, respectively). The mutant channel model (Fig. 1b) contains two possible modes of gating, a ‘background (dispersed) mode’ and a ‘burst mode’. The background mode includes the above six states (Fig. 1b); it is similar to the wild-type model except for alterations in the voltage dependence of activation, inactivation and recovery from inactivation (Box 1). Most (>99%) of the mutant channels reside in the background mode states. The models were incorporated into the Luo–Rudy model (Fig. 1c) for action potential simulations.

Box 1 Simulation methods

The general approach to modelling the action potential is the same as that described for the Luo–Rudy model^{8–10} except that the I_{Na} transmembrane current is reformulated from the single-channel kinetics. We use the general approach of refs 20 and 21. All kinetic parameters were normalized to 37 °C with a Q_{10} of 3 (ref. 19).

All the simulations were encoded in C/C++. Simulations were implemented (double precision) on a Sun Workstation Ultra 1. A time step of 0.005 ms was used during the stimulus and the action potential upstroke. At all other times, a 0.01-ms time step was used.

Transition rates

Wild-type channel (ms^{-1}):

$$\begin{aligned} C3 \rightarrow C2 & \alpha_{11} = 3.802/(0.1027 \times \exp(-v/17.0) + 0.20 \times \exp(v/150)) \\ C2 \rightarrow C1 & \alpha_{12} = (3.802/(0.1027 \times \exp(-v/15.0) + 0.23 \times \exp(-v/150))) \\ C1 \rightarrow O & \alpha_{13} = 3.802/(0.1027 \times \exp(-v/12.0) + 0.250 \times \exp(-v/150)) \\ C2 \rightarrow C3 & \beta_{11} = 0.1917 \times \exp(-v/20.3) \\ C1 \rightarrow C2 & \beta_{12} = 0.20 \times \exp(-(v-5)/20.3) \\ O \rightarrow C1 & \beta_{13} = 0.22 \times \exp(-(v-10)/20.3) \\ O \rightarrow IF & \alpha_2 = (9.178 \times \exp(v/29.68)) \\ IF \rightarrow O & \beta_2 = ((\alpha_{13} \times \alpha_2 \times \alpha_3)/(\beta_{13} \times \beta_3)) \\ IF \rightarrow C1 & \alpha_3 = (3.7933^{-9} \times \exp(-v/5.2)) \\ C1 \rightarrow IF & \beta_3 = (0.0084 + 0.00002 \times v) \\ IF \rightarrow IS & \alpha_4 = \alpha_2/100 \\ IS \rightarrow IF & \beta_4 = \alpha_3 \end{aligned}$$

Δ KPQ mutant channel* (ms^{-1}):

$$\begin{aligned} xC3 \rightarrow xC2 & \alpha_{11} = 1.25 \times (3.082/(0.1027 \times \exp(-v/17.0) + 0.20 \times \exp(-v/150))) \\ xC2 \rightarrow xC1 & \alpha_{12} = 1.25 \times (3.082/(0.1027 \times \exp(-v/15.0) + 0.23 \times \exp(-v/150))) \\ xC1 \rightarrow xO & \alpha_{13} = 1.25 \times (3.082/(0.1027 \times \exp(-v/12.0) + 0.250 \times \exp(-v/150))) \\ xC2 \rightarrow xC3 & \beta_{11} = 0.1917 \times \exp(-v/20.3) \\ xC1 \rightarrow xC2 & \beta_{12} = 0.20 \times \exp(-(v-5)/20.3) \\ xO \rightarrow xC1 & \beta_{13} = 0.22 \times \exp(-(v-10)/20.3) \\ O \rightarrow IF & \alpha_2 = (9.178 \times \exp(v/100)) \\ IF \rightarrow O & \beta_2 = ((\alpha_{13} \times \alpha_2 \times \alpha_3)/(\beta_{13} \times \beta_3)) \\ IF \rightarrow UC1 & \alpha_3 = 20 \times (3.7933^{-9} \times \exp(-v/5.2)) \\ UC1 \rightarrow IF & \beta_3 = 2 \times (0.0084 + 0.00002 \times v) \\ IF \rightarrow IS & \alpha_4 = \alpha_2/100 \\ IS \rightarrow IF & \beta_4 = \alpha_3 \end{aligned}$$

* x represents U or L, as transition rates in the background or burst modes are the same.

Transition rates between modes are background to burst, $\mu_1 = 2 \times 10^{-6} \text{ ms}^{-1}$; burst to background, $\mu_2 = 1 \times 10^{-4} \text{ ms}^{-1}$.

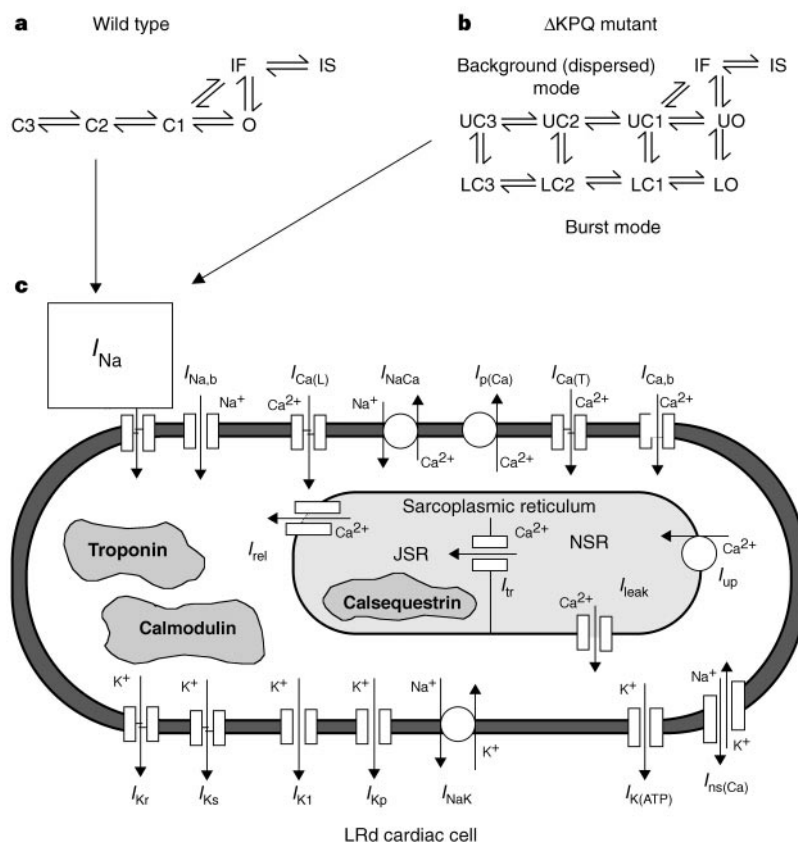


Figure 1 Markovian models of sodium channels. **a**, Wild-type (WT); **b**, Δ KPQ mutant. **c**, The Luo-Rudy dynamic cardiac-cell model (LRd)⁸⁻¹⁰. JSR, junctional sarcoplasmic reticulum; NSR, network sarcoplasmic reticulum.

Figure 2a shows simulated and experimentally measured⁶ single-channel recordings. Like wild-type channels (Fig. 2a, WT), mutant channels in the background mode open and inactivate in response to depolarization (Fig. 2a, Δ KPQ, top three traces). Mutant channels activate more quickly (increased rate constants of activation). As the open and the inactivated states are strongly coupled, the increased probability of being in the open state leads to increased probability of inactivation and faster decay of the mutant macroscopic current compared to the wild type^{6,7}. Mutant channels also recover from inactivation more rapidly than wild-type channels¹¹. As channel availability is increased, a background current results from channel reopenings at depolarized membrane potentials (late openings following the first opening in Fig. 2a, Δ KPQ; top three traces). Thus, an increased rate of recovery from inactivation in the background mode leads to dispersed channel re-openings and a late component of I_{Na} .

The burst mode of the Δ KPQ channel model (Fig. 1b) does not include the inactivated states IF and IS. Being in this mode of the mutant channel corresponds to transient failure of the channel to inactivate (Fig. 1b). The likelihood of entry into the lower states of the Δ KPQ burst mode is very low ($2 \times 10^{-6} \text{ ms}^{-1}$) but once channels are in these states, return to the background mode is unlikely (10^{-4} ms^{-1}). This results in a small population of channels in the lower states that fail to inactivate. These channels bounce back and forth between closed available states and a single open state (LC1, LC2 and LC3; and LO, respectively) (Fig. 2a, Δ KPQ; bottom two traces). Note the frequent opening events ('bursting') of channels in this mode. Together with the dispersed re-openings of channels in the background mode, this behaviour can result in a significant late current that causes marked changes in the action potential.

The mean open time (MOT) of simulated and measured Δ KPQ mutant channels is shown in Fig. 2b, with background mode in the

top panels and burst mode in the bottom panels. The experiments show relatively weak voltage dependence for the MOT of background openings, as compared to strong voltage dependence for the MOT of burst-type openings⁶. This behaviour is reflected in the similarity of the MOT distributions for depolarization to -50 mV or -20 mV in the background mode, in contrast to the widening of the distribution in the burst mode at -20 mV compared to -50 mV . In our simulations, the voltage dependence of the MOT is reproduced by simply removing inactivation from the background mode (that is, being in the burst mode of the mutant channel model). In the absence of inactivation, the voltage dependence of bursting is dominated by deactivation (Fig. 1b, transition from LO to LC3). As the membrane potential is progressively depolarized, the deactivation transition becomes slower and the mean open time increases ($1/\beta_{13}$). In contrast, the weak voltage dependence of the background mode MOT ($1/(\beta_{13} + \alpha_2)$) in this voltage range is due to the co-dominance of deactivation (transition from UO to UC3) and inactivation (transition from UO to IF)¹². Increased depolarization results in a decrease in the deactivation transition rate that is balanced by a concomitant increase in the inactivation rate. This results in little dependence of the MOT on voltage in this range.

The additive effects of background channel reopenings and a small subpopulation of bursting channels on the macroscopic current are shown in Fig. 3a. In the wild-type channels, depolarization to -20 mV from a holding potential of -140 mV results in a large macroscopic current (peak $> 300 \mu\text{A } \mu\text{F}^{-1}$) that fully inactivates and returns to zero. In contrast, a model cell with a 100% Δ KPQ mutant-channel population is marked by a persistent component ($0.3 \mu\text{A } \mu\text{F}^{-1}$) of current even after prolonged depolarization (200 ms). Experimental data are shown for comparison (right panels)⁷.

Figure 3b shows the accelerated decay of the early component of macroscopic sodium current in the Δ KPQ mutant, as seen

experimentally^{6,7}. The cell is depolarized to -40 mV from a holding potential of -140 mV (inset). The increase in the rate of current decay results largely from an increased rate of activation. Faster activation (increased probability in UO) leads to an increased probability of inactivation and corresponding current decay. We compare simulated data to experimental data from ref. 6.

So far, we have established the capability of the Markov sodium-channel model to correctly reproduce single-channel behaviour and macroscopic current properties. We next investigated the cellular arrhythmogenic consequences of the Δ KPQ mutation by integrating the channel models into the Luo–Rudy model of the cardiac ventricular cell^{8–10} (Fig. 1).

Figure 4A and B shows the action potential and sodium current (I_{Na}) from the tenth beat of a simulated cell, paced at a cycle length of 400 ms. In Fig. 4A, the cell contains 100% wild-type channels; in Fig. 4B the channels are Δ KPQ mutants. The mutant-cell action potential (Fig. 4B) has an increased depolarization rate compared to the wild type (325.9 and 296.9 V s⁻¹, respectively), as well as marked prolongation of the action potential duration (APD; 62.3 ms increase). Inspection of the corresponding sodium current at high gain (Fig. 4B, bottom) reveals a persistent component of inward current ($3 \mu\text{A } \mu\text{F}^{-1}$) during the plateau of the action potential. The morphology of the simulated current (Fig. 4B) compares well with experimentally recorded I_{Na} during an action potential clamp (Fig. 4D)¹³. Although the amplitude of the persistent current is very small, only 1.0% of the peak I_{Na} ($>300 \mu\text{A } \mu\text{F}^{-1}$), it is sufficient to shift the delicate balance of currents that exists during the action potential plateau^{8–10,14}. The additional inward current carried by sodium acts in opposition to repolarizing outward currents and

prolongs the APD. Figure 4C shows the effect of decreasing the pacing rate (cycle length from 400 ms to 600 ms) in the Δ KPQ mutant cell. The tenth beat is shown and demonstrates an arrhythmogenic early after-depolarization (EAD) that develops from plateau potentials^{9,14}. At the slower rate, repolarizing outward currents (in particular the slow delayed rectifier, I_{Ks}) deactivate more completely between beats^{10,14}. The combination of the persistent depolarizing I_{Na} and the reduced repolarizing I_{Ks} causes marked APD prolongation, which gives time for the L-type calcium current, $I_{Ca(L)}$, to recover from inactivation and reactivate. The secondary reactivation of $I_{Ca(L)}$ (not shown) depolarizes the membrane and generates the EAD¹⁴. A small population of bursting Δ KPQ channels ($P(\text{LO}) < 0.005$), in combination with dispersed channel re-openings ($P(\text{UO}) < 0.001$), contributes to the persistent inward current.

Transcription of both mutant and wild-type SCN5A alleles is likely to occur, as affected individuals are typically heterozygous for LQT3. To investigate this possibility, we simulated paced cells containing different WT: Δ KPQ ratios of Na⁺ channels. Figure 4E shows the effects of varying the stoichiometry of wild-type and Δ KPQ channel populations in the cell. Action potentials from a wild-type cell paced for five beats at a cycle length of 1,000 ms are shown in Fig. 4E, a. As the percentage of Δ KPQ channels is increased relative to wild type (Fig. 4E, b–c), the APD becomes progressively longer and EAD development is marked. Note (Fig. 4E, b) that the presence of 50% mutant channels is sufficient to cause EADs at this rate (cycle length, 1,000 ms).

The development of arrhythmogenic episodes in congenital LQT3 is correlated with bradycardia during sleep or relaxation¹⁵. Faster pacing rates result in adaptive shortening of the APD and an

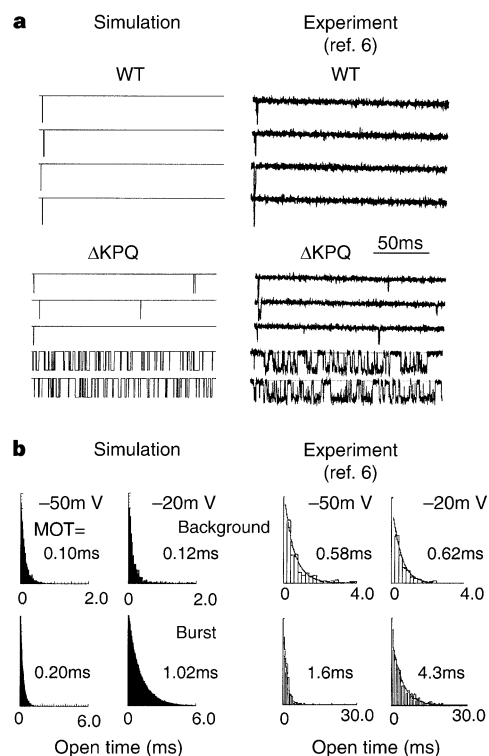


Figure 2 Single-channel properties of WT and mutant I_{Na} . **a**, Single-channel gating in simulated channels (left) and experimental recordings from ref. 6 (right). Top: wild-type (WT) channel gating. Bottom Δ KPQ mutant channel gating, where top three traces correspond to the mutant channel in the background mode and bottom two traces show burst opening in the burst mode. **b**, The MOT of Δ KPQ mutant channels. Simulations are compared to data from ref. 6. Differences in MOT are due to temperature differences and can be compensated for by a Q_{10} of 3 (ref. 19).

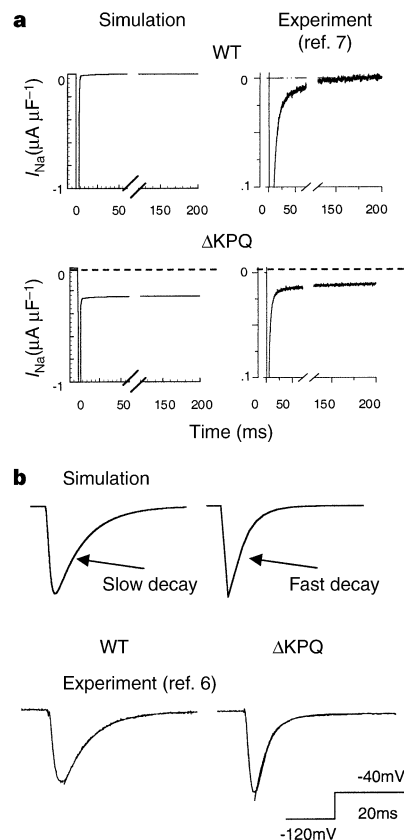


Figure 3 Macroscopic WT and mutant I_{Na} . **a**, Macroscopic I_{Na} during step depolarization from -140 to -20 mV. Top: Wild-type channels. I_{Na} inactivates and returns to zero. Bottom: Δ KPQ channels. There is a persistent component of I_{Na} . Simulations (left) are compared to normalized data from ref. 7 (right). **b**, Increased rate of macroscopic current decay in Δ KPQ mutant channels. Simulations (top) are compared to data from ref. 6 (bottom). Protocol is shown in inset.

accompanying shortening of the Q–T interval in the ECG. To investigate the cellular response to changes in cycle length, we chose a WT:ΔKPQ ratio of 50:50. Figure 4F shows the response of the cell to bradycardia, tachycardia and normal physiological cycle length. Figure 4F, a shows the fourth and fifth beats at rapid pacing (cycle length, 400 ms). Note the absence of EADs and adaptive shortening of the action potential duration at faster cycle length. At a normal human cycle length (Fig. 4F, b), the action potential duration is markedly prolonged relative to the wild type. When the cell is exposed to bradycardia (cycle length, 1,200 ms; Fig. 4F, c), hallmark EADs develop after several beats and are sustained during subsequent pacing.

This study describes the effects of a genetic defect (ΔKPQ mutation) on the cardiac ventricular action potential and demonstrates its arrhythmogenic consequences at the cellular level. We have formulated a Markovian based model that represents discrete structural states and their interactions. Although the traditional Hodgkin–Huxley¹⁶ type of model is useful for a variety of modelling efforts, this framework is insufficient for incorporation of state-specific structural channel defects. Strong coupling between discrete structural states (a property not represented in the Hodgkin–Huxley scheme) is required, as a mutation affecting the structural integrity of a particular state will affect adjacent states^{17,18}. The theoretical tool introduced here allows the integration of single channels (wild-type or mutant) into the whole cell, where their effects on cellular behaviour can be studied. As more idiopathic diseases are linked to congenital abnormalities, modelling strategies of this type can help to bridge the gap between genetic molecular defects and their phenotypic consequences. □

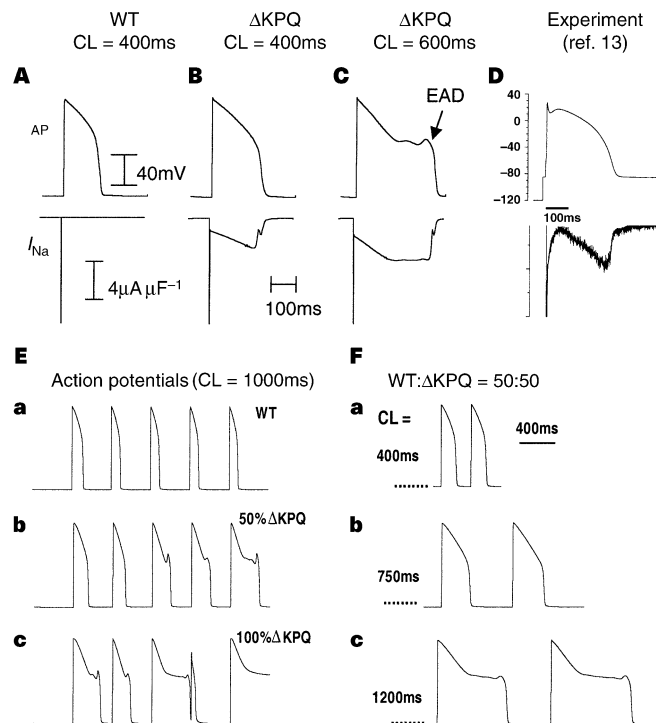


Figure 4 Effect of ΔKPQ mutation on the whole-cell action potential. **A–D**, Persistent I_{Na} during the action potential plateau (**B**) prolongs the action potential duration (compare to WT in **A**). Note close correspondence of the morphology of I_{Na} between simulation (**B**) and data from ref. 13 (**D**). Slowing the rate (**C**) results in further APD prolongation and EAD development. **E**, The effects of varying the WT:ΔKPQ channel stoichiometry within a single cell. The cell is paced at a cycle length of 1,000 ms for five beats. 50% mutant channels are sufficient to cause EADs. **F**, Effect of pacing cycle length on EAD development. Cells contain 50% mutant channel population and are paced for five beats at the indicated cycle length (CL) (4th and 5th beats are shown).

Received 23 March; accepted 8 June 1999.

- Welsh, M. J. & Hoshi, T. Ion channels lose the rhythm. *Nature* **376**, 640 (1995).
- Attali, B. A new wave for heart rhythms. *Nature* **384**, 24–25 (1996).
- Keating, M. T. & Sanguinetti, M. C. Molecular genetic insights into cardiovascular disease. *Science* **272**, 681–685 (1996).
- Barinaga, M. Tracking down mutations that can stop the heart. *Science* **281**, 32–34 (1998).
- Dumaine, R. *et al.* Multiple mechanisms of Na⁺ channel-linked long-QT syndrome. *Circ. Res.* **78**, 916–924 (1996).
- Chandra, R., Starmer, C. F. & Grant, A. O. Multiple effects of the KPQ deletion on gating of human cardiac Na⁺ channels expressed in mammalian cells. *Am. J. Physiol.* **274**, H1643–H1654 (1998).
- Bennett, P. B., Yazawa, K., Makita, N. & George, A. L. Jr Molecular mechanism for an inherited cardiac arrhythmia. *Nature* **376**, 683–685 (1995).
- Luo, C. H. & Rudy, Y. A dynamic model of the cardiac ventricular action potential. I. Simulations of ionic currents and concentration changes. *Circ. Res.* **74**, 1071–1096 (1994).
- Luo, C. H. & Rudy, Y. A dynamic model of the cardiac ventricular action potential. II. After-depolarizations, triggered activity, and potentiation. *Circ. Res.* **74**, 1097–1113 (1994).
- Zeng, J., Laurita, K. R., Rosenbaum, D. S. & Rudy, Y. Two components of the delayed rectifier K⁺ current in ventricular myocytes of the guinea pig type. Theoretical formulation and their role in repolarization. *Circ. Res.* **77**, 140–152 (1995).
- Wang, D. W., Yazawa, K., George, A. L. Jr & Bennett, P. B. Characterization of human cardiac Na⁺ channel mutations in the congenital long QT syndrome. *Proc. Natl Acad. Sci. USA* **93**, 13200–13205 (1996).
- Yue, D. T., Lawrence, J. H. & Marban, E. Two molecular transitions influence cardiac sodium channel gating. *Science* **244**, 349–352 (1989).
- Wang, D. W., Yazawa, K., Makita, N., George, A. L. Jr & Bennett, P. B. Pharmacological targeting of long QT mutant sodium channels. *J. Clin. Invest.* **99**, 1714–1720 (1997).
- Zeng, J. & Rudy, Y. Early afterdepolarizations in cardiac myocytes: mechanism and rate dependence. *Biophys. J.* **68**, 949–964 (1995).
- Schwartz, P. J. *et al.* Long QT syndrome patients with mutations of the SCN5A and HERG genes have differential responses to Na⁺ channel blockade and to increases in heart rate. Implications for gene-specific therapy. *Circulation* **92**, 3381–3386 (1995).
- Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500–544 (1952).
- Tomaselli, G. F. *et al.* A mutation in the pore of the sodium channel alters gating. *Biophys. J.* **68**, 1814–1827 (1995).
- Mitsuiye, T. & Noma, A. Inactivation of the cardiac Na⁺ channels in guinea-pig ventricular cells through the open state. *J. Physiol.* **485**, 581–594 (1995); erratum *J. Physiol. (Lond.)* **488**, 841 (1995).
- Hanck, D. A. in *Cardiac Electrophysiology: From Cell to Bedside* (eds Zipes, W. & Jalife, J.) 65–74 (W. B. Saunders, Philadelphia, 1995).
- Balser, J. R., Roden, D. M. & Bennett, P. B. Global parameter optimization for cardiac potassium channel gating models. *Biophys. J.* **57**, 433–444 (1990).
- Zeng, J. in *Biomedical Engineering* (Case Western Reserve Univ., Cleveland, 1997).

Acknowledgements. We thank P. Bennett and A. Grant for providing us with original figures from their experimental publications. This work was supported by the NIH-National Heart, Lung and Blood Institute and The Whitaker Foundation.

Correspondence and requests for materials should be addressed to Y.R. (e-mail: yxr@po.cwru.edu).

Developmental and activity-dependent regulation of kainate receptors at thalamocortical synapses

Fleur L. Kidd & John T. R. Isaac

MRC Centre for Synaptic Plasticity, Department of Anatomy, University of Bristol, Bristol BS8 1TD, UK

Most of the fast excitatory synaptic transmission in the mammalian brain is mediated by ionotropic glutamate receptors, of which there are three subtypes: AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate), NMDA (N-methyl-D-aspartate) and kainate. Although kainate-receptor subunits (GluR5–7, KA1 and 2) are widely expressed in the mammalian central nervous system^{1,2}, little is known about their function. The development of pharmacological agents that distinguish between AMPA and kainate receptors has now allowed the functions of kainate receptors to be investigated^{3–7}. The modulation of synaptic transmission by kainate receptors^{8–14} and their synaptic activation^{8–14} in a variety of brain regions have been reported. The expression of kainate receptor subunits is developmentally regulated^{1,2} but their role in plasticity and development is unknown. Here we show that developing thalamocortical synapses express postsynaptic kainate receptors as well as AMPA receptors; however, the two receptor subtypes do not co-localize. During the critical period for experience-dependent plasticity, the kainate-receptor contribution to transmission decreases; a similar decrease occurs when long-term potentiation

is induced *in vitro*. This indicates that during development there is activity-dependent regulation of the expression of kainate receptors at thalamocortical synapses.

In the presence of the NMDA-receptor (NMDAR) antagonist D-2-amino-5-phosphonopentanoate acid (D-AP5; 100 μ M) and the GABA_A-receptor antagonist picrotoxin (50 μ M), stimulation of thalamocortical axons evoked a dual-component excitatory postsynaptic current (EPSC) in layer-IV neurons of the primary somatosensory (barrel) cortex of neonatal (postnatal day (P) 3–8) rats (Fig. 1a). The EPSC decays were best fitted by the sum of two exponentials ($\tau_{\text{fast}} = 5.9 \pm 0.5$ ms, $\tau_{\text{slow}} = 169.2 \pm 16.7$ ms, $n = 93$). The fast component of the EPSC was mediated by AMPARs, as indicated by its sensitivity to the selective AMPAR antagonist GYKI 53655 (1-(4-aminophenyl)-3-methylcarbamyl-4-methyl-7, 8-methylenedioxy-3, 4-dihydro-5H-2, 3-benzodiazepine; 25 μ M active isomer; refs 15, 16). The decay of the GYKI-resistant component was best fitted by a single exponential similar to τ_{slow} of the dual EPSC ($\tau_{\text{GYKI}} = 154.9 \pm 25.5$ ms, $n = 10$, $P = 0.85$ versus τ_{slow} ; Fig. 1a, b). The blockade of the GYKI-resistant component by the AMPA/kainate-receptor antagonist 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX; 50–100 μ M, $n = 5$; Fig. 1a) indicates that this component was mediated by kainate receptors (KARs). The pharmacological profile and kinetics of this mixed AMPA and kainate EPSC is similar to those reported for synapses on CA1 hippocampal interneurons^{10,11}.

To characterize the KARs further, we studied the current–voltage (*I*–*V*) relationship of the EPSCs. Measured at the peak, the *I*–*V* relationship was approximately linear with a reversal potential of ~ 0 mV, as expected for a conductance mediated by AMPARs containing edited GluR2 subunits¹ (Fig. 1c). At the tail, however, the EPSC exhibited a strongly rectifying *I*–*V* relationship (Fig. 1d), as did the pharmacologically isolated kainate EPSC (data not shown). The mean rectification index of the KAR-mediated EPSCs (ratio of estimated conductance at +40 mV and –60 mV) was 0.13 ± 0.08 ($n = 24$). This is characteristic of a neonatal KAR predominantly containing unedited GluR5 and/or GluR6 subunits, which is blocked in a voltage-dependent manner by polyamines^{17–19}. The low rectification index indicates that these channels are significantly permeable to calcium ions^{20,21}.

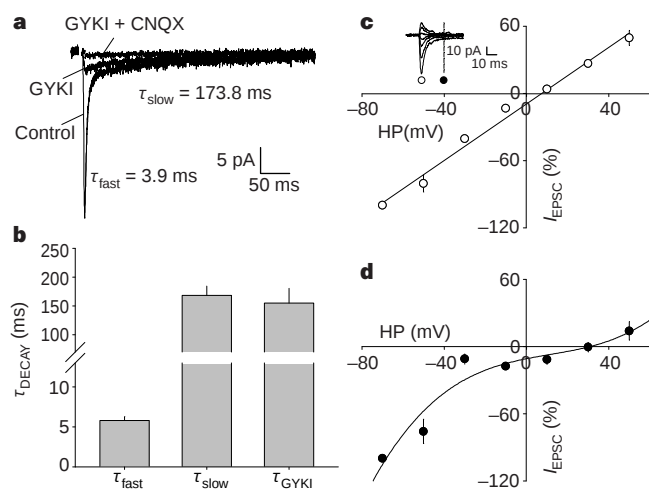


Figure 1 KAR-mediated synaptic transmission at developing thalamocortical synapses. **a**, Superimposed EPSCs in: 'Control', D-AP5 (100 μ M), picrotoxin (50 μ M); 'GYKI' D-AP5, picrotoxin, GYKI 53655 (25 μ M active isomer); 'GYKI + CNQX' D-AP5, picrotoxin, GYKI 53655, CNQX (100 μ M). For this and the following figures, thick line represents best exponential fit (single or double). **b**, Mean decay time constants for dual component (τ_{fast} and τ_{slow} , $n = 93$) and GYKI-resistant EPSCs (τ_{GYKI} , $n = 10$). **c, d**, Summary *I*–*V* analysis of peak (**c**) and tail (**d**) of dual-component EPSCs ($n = 18$). Continuous lines represent linear regressions. Inset in **c**: single example of *I*–*V* series. HP, holding potential.

Data from the investigation of the dual-component EPSCs suggest that thalamocortical synapses contain both AMPA and kainate receptors. However, in some neurons kainate EPSCs could be evoked in the absence of an AMPA component (kainate-only EPSCs; $n = 14$; Fig. 2a). These EPSCs decayed with a single exponential with a similar time constant to τ_{slow} and τ_{GYKI} ($\tau_{\text{KA-only}} = 194.5 \pm 25.2$ ms, $n = 14$, $P = 0.49$ versus τ_{slow} , $P = 0.29$ versus τ_{GYKI} ; Fig. 2b). The *I*–*V* relationship of kainate-only EPSCs showed that they also rectified strongly (Fig. 2c). In other neurons, AMPA EPSCs could be evoked in the absence of a kainate component (AMPA-only EPSCs; $n = 14$; Fig. 2d). These decayed with a single time constant very similar to τ_{fast} of the dual component EPSC ($\tau_{\text{AMPA-only}} = 5.3 \pm 0.6$, $n = 14$, $P = 0.71$ versus τ_{fast} ; Fig. 2e). In a further subset of neurons, a mixture of dual component, AMPA-only and kainate-only EPSCs was observed in response to stimuli of the same intensity (Fig. 2f). In these cases, all three types of EPSC had the same threshold for activation and identical latencies, indicating that they resulted from the activation of the same population of axons. Analysis of all cells showed that there was no difference in the latencies of the three types of synaptic response (ventrobasal stimulation: dual EPSC was 11.2 ± 0.4 ms, $n = 23$; kainate only was 11.4 ± 2.0 ms, $n = 4$; AMPA only was 11.8 ± 0.5 ms, $n = 5$; white-matter stimulation: dual EPSC was 5.2 ± 0.2 ms, $n = 77$; kainate only was 5.9 ± 0.8 ms, $n = 10$; AMPA only was 5.9 ± 0.5 , $n = 9$; no significant difference between any of these data within the ventrobasal or white-matter groups).

This characterization of evoked EPSCs suggests that some developing thalamocortical synapses have KARs but no AMPARs ('kainate synapses'), whereas others have AMPARs but no KARs ('AMPA synapses'). This raises the question of whether KARs and AMPARs ever co-localize at thalamocortical synapses. To address this, we analysed spontaneous EPSCs (SEPSCs) that were collected simultaneously with the evoked currents. When fast-rising SEPSCs were aligned and averaged, no slow KAR-mediated tail was evident

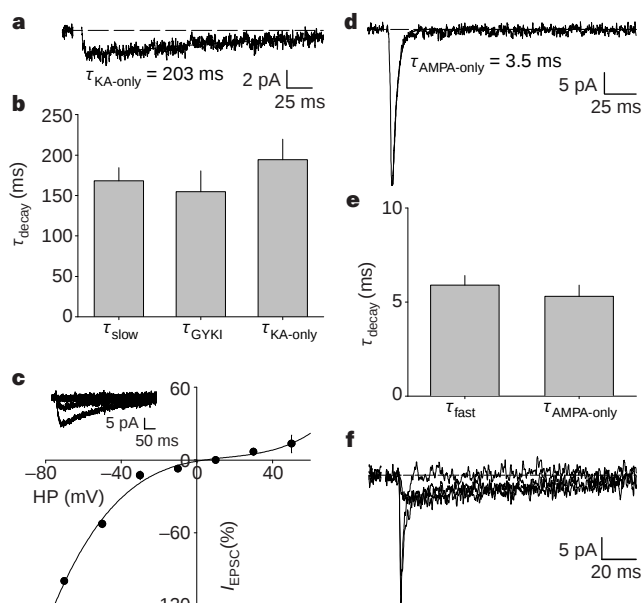


Figure 2 KAR-mediated EPSCs can be evoked in the absence of an AMPAR component. **a**, Example of a kainate EPSC with no AMPAR component. **b**, Summary of kainate EPSC decay time-constants ($\tau_{\text{KA-only}}$; $n = 14$; τ_{slow} and τ_{GYKI} replotted). **c**, Kainate EPSC *I*–*V* summary data ($n = 7$; continuous line is linear regression). Inset: example traces from one cell. **d**, Example of an AMPA EPSC with no KAR component (white-matter stimulation). **e**, Summary of AMPA EPSC decay time constants ($\tau_{\text{AMPA-only}}$; $n = 14$; τ_{fast} re-plotted). **f**, AMPA-only, dual component and kainate-only EPSCs (superimposed) from one cell (same stimulus intensity; ventrobasal stimulation).

even though the simultaneously collected evoked EPSC exhibited both components (Fig. 3a–c). The decays of these SEPSCs were always best fitted with a single exponential similar to τ_{fast} and $\tau_{\text{AMPA-only}}$ ($\tau_{\text{SEPSC}} = 4.3 \pm 0.5$ ms, $n = 15$, $P = 0.27$ versus τ_{fast} , $P = 0.24$ versus $\tau_{\text{AMPA-only}}$; Fig. 3d). Although the source of the SEPSCs in these neurons is likely to be both thalamocortical and intracortical synapses, if AMPARs and KARs coexist at thalamocortical synapses, a proportion of the SEPSCs would be expected to exhibit both components.

These data indicate that AMPARs are not co-localized with KARs at thalamocortical synapses. One prediction from this is that a proportion of SEPSCs should be mediated by KARs alone, representing spontaneous release at kainate synapses. Slow spontaneous EPSCs (slow SEPSCs) were found to be present simultaneously with fast SEPSCs within the same neurons (Fig. 4a). Slow SEPSCs had an almost identical time course to pharmacologically isolated evoked kainate EPSCs (Fig. 4b), persisted in the presence of GYKI 53655 (25 μ M active isomer, $n = 5$; data not shown) but were blocked by CNQX (50–100 μ M, $n = 4$; data not shown). Slow SEPSCs decayed with a single time constant similar to that for evoked kainate EPSCs ($\tau_{\text{slow SEPSC}} = 139.1 \pm 20.5$, $n = 7$, $P = 0.67$ versus τ_{slow} , $P = 0.66$ versus τ_{GYKI} , $P = 0.17$ versus $\tau_{\text{KA-only}}$; Fig. 4c). Thus slow SEPSCs displayed identical pharmacological and kinetic properties to evoked kainate EPSCs, indicating that they were due to spontaneous glutamate release at kainate synapses.

The characterization of the evoked and SEPSCs indicates that two types of thalamocortical input exist in the developing barrel cortex, kainate synapses and AMPA synapses, and that the two types of receptor do not coexist at the same synapses. The ages of the animals used in this study coincide with the critical period for experience-dependent plasticity at this input²². Therefore we investigated whether the relative amount of charge contributed by the AMPAR- and KAR-mediated components changed during this developmental period. The contribution of KARs was large early in development (P3–5) but progressively decreased from P6 to P8 (Fig. 5a). This was not due to a change in KAR kinetics (Fig. 5b) but instead to a decrease in the number of kainate synapses contributing to the EPSC (Fig. 5c). This was confirmed by the decrease in the number of pathways that exhibited kainate-only EPSCs and the coincident increase in the number of AMPA-only pathways that occurred during this developmental period (Fig. 5d).

Experience-dependent plasticity at thalamocortical synapses in the barrel cortex is activity- and NMDAR-dependent²² and correlates with the developmental profile of long-term potentiation

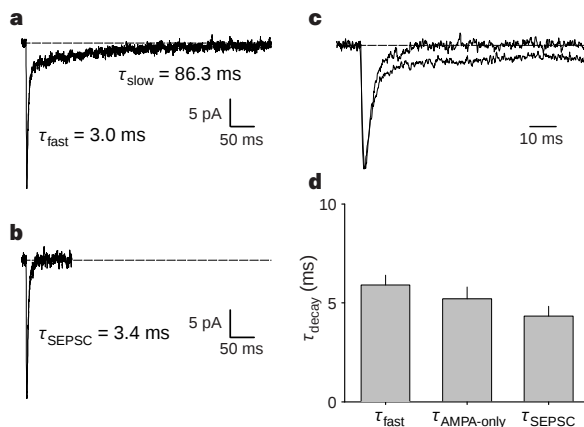


Figure 3 AMPAR-mediated spontaneous EPSCs do not have a KAR component. Examples of **a**, averaged evoked EPSC; **b**, averaged fast SEPSC (same cell, collected during same 5 min period); and **c**, same traces superimposed (scaled; expanded timescale; thick line is SEPSC). **d**, Mean decay time constant for fast SEPSCs (τ_{SEPSC} ; $n = 15$; τ_{fast} and $\tau_{\text{AMPA-only}}$ replotted).

(LTP)^{22–24}. Therefore we investigated whether kainate synapses could be converted to AMPA synapses during LTP. In slices from P4–P7 animals, pairing at 0 mV resulted in the induction of LTP^{23–25} that lasted for the duration of the recording (Fig. 6a). This was associated with a decrease in the charge carried by the KAR component of the EPSC ($62 \pm 11\%$ of baseline, $P < 0.05$, $n = 11$; Fig. 6b, c). In two neurons, LTP caused the kainate component to disappear completely so that the decay of the potentiated EPSC was best fitted by a single fast exponential (data not shown). This reduction in the KAR-mediated component was dependent upon LTP induction; there was no change in the KAR-mediated component in interleaved controls in which early and late portions of an extended baseline were compared ($n = 3$) or in which the pairing protocol failed to induce LTP ($n = 6$), presumably due to ‘wash-out’ (KAR component was $98 \pm 26\%$ of baseline, $n = 9$; Fig. 6d). One important assumption underlying these experiments is that, in the absence of D-AP5, at the holding potential of -70 mV, the slow component of the EPSC was not significantly contaminated by an NMDAR-mediated current. τ_{decay} under these conditions was similar to that observed in the presence of D-AP5 ($\tau_{\text{decay}} = 184.5 \pm 28.1$, $n = 25$) and, in control experiments, addition of D-AP5 (25–100 μ M) had no effect on the proportion of the dual component EPSC that was mediated by the slow component (slow component was $102 \pm 21\%$ of baseline, $n = 5$; Fig. 6d). This confirmed that

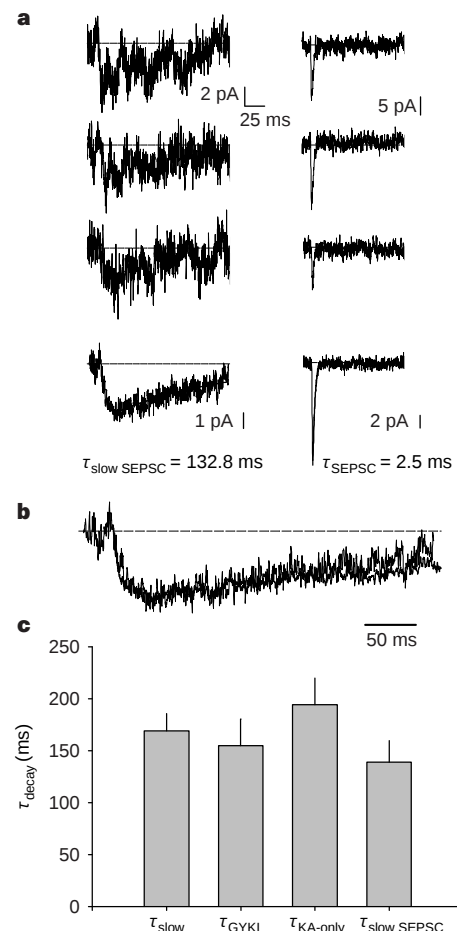


Figure 4 Slow spontaneous EPSCs with similar kinetics and pharmacology to evoked kainate EPSCs coexist with AMPAR-mediated SEPSCs. **a**, Single examples (top three traces) and average (bottom trace) for slow (left) and fast (right) SEPSCs collected from the same cell during the same time period. **b**, Averaged slow SEPSC (from **a**) and an evoked kainate EPSC (in GYKI, thick line; scaled, superimposed). **c**, Mean decay time constant for all slow SEPSCs ($\tau_{\text{slow SEPSC}}$; $n = 7$; τ_{slow} , τ_{GYKI} and $\tau_{\text{KA-only}}$ replotted).

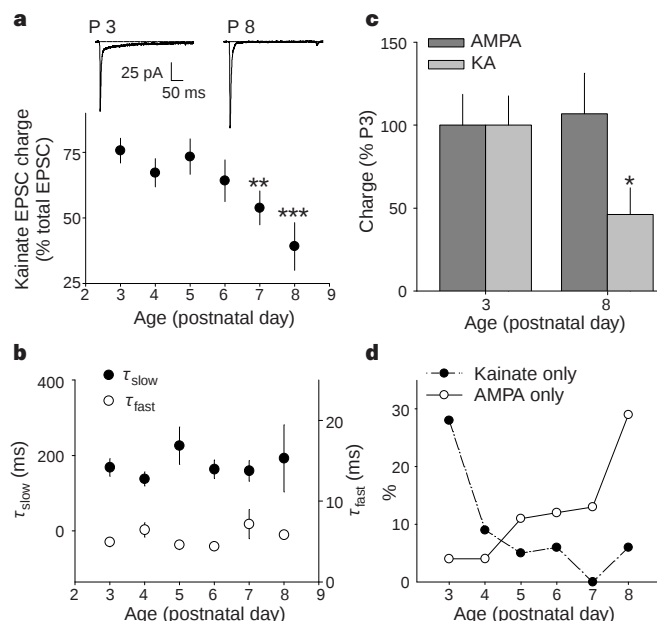


Figure 5 The KAR component of thalamocortical transmission decreases during the critical period. **a**, Developmental profile of the KAR component of the dual EPSC (P3 $n = 25$, P4 $n = 23$, P5 $n = 19$, P6 $n = 17$, P7 $n = 23$, P8 $n = 17$; *** $P < 0.0005$ P8 versus P3; ** $P < 0.01$ P7 versus P3). Insets: example EPSCs (scale bar applies to both). **b**, τ_{slow} and τ_{fast} versus age. **c**, Charge through AMPAR and KAR components of the EPSC (% of charge at P3; * $P < 0.05$). **d**, Number of kainate-only or AMPA-only pathways (% total pathways) versus age.

there was no detectable contribution from NMDARs to the KAR component of transmission under these experimental conditions.

The correlation between the developmental profile of KAR-mediated transmission and the critical period for experience-dependent plasticity at thalamocortical inputs indicates that kainate synapses are modified in an activity-dependent manner during development. The finding that LTP causes a rapid decrease in the KAR-mediated EPSC provides direct evidence for such a process. One possible mechanism for this could be the rapid conversion of kainate synapses to AMPA synapses, in which KARs are rapidly removed and replaced by AMPARs in a highly coordinated manner. This is an extension of the mechanism proposed for the conversion of silent to functional synapses at thalamocortical synapses during the critical period²⁴. Other possible mechanisms include a rapid biochemical modification of KARs causing a large reduction in channel conductance or permanently desensitizing receptors. A mechanism involving a change in subunit composition is improbable because of the rapidity of the loss of the kainate response. A spill-over mechanism, as has been proposed for LTP at silent synapses²⁶, is unlikely because the properties of KARs in this study indicate that they are composed of low-affinity subunits^{4,17,27} resulting in a receptor complex that has a low affinity for L-glutamate like AMPARs²⁸. Kainate synapses could confer some novel features on neonatal thalamocortical circuitry. The slow kinetics of the kainate EPSC indicates that, at hyperpolarized potentials, kainate synapses will provide a relatively long-lasting depolarization of layer-IV neurons in response to low-frequency transmission, and this may be associated with an influx of calcium ions. However, depolarization will prevent transmission because of the voltage-dependent block by spermine. The polyamine block also provides a potential mechanism for short-term plasticity owing to its activity-dependent relief²⁹. Thus the presence of KARs at thalamocortical synapses and the developmental conversion of transmission from predominantly KAR-mediated to AMPAR-mediated during the critical period could have important consequences for the mechanisms by which experience modifies cortical circuitry during development. □

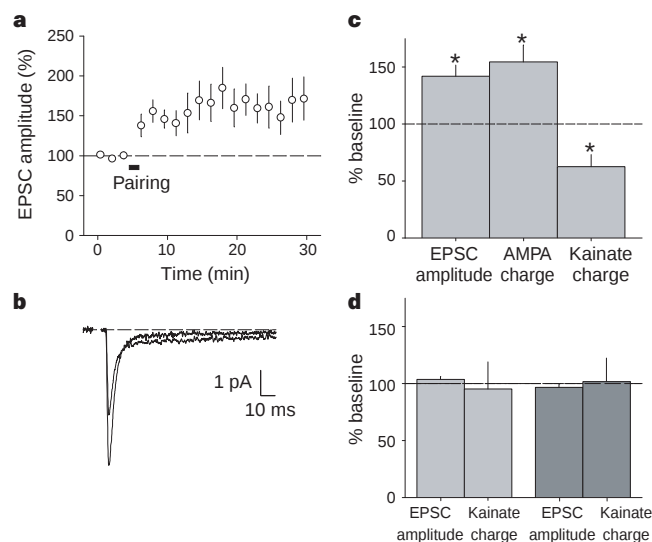


Figure 6 LTP induction causes a rapid reduction in the KAR component of EPSCs. **a**, Amplitude versus time (% baseline) for all LTP experiments ($n = 11$). **b**, Average EPSCs (superimposed; average of all responses) for baseline (thick line) and LTP (thin line), for one experiment. **c**, Summary of EPSC amplitude, AMPAR and KAR charge for LTP experiments (% baseline; $n = 11$; * $P < 0.05$). **d**, EPSC amplitude, AMPAR and KAR charge for (light bars) wash-out LTP and extended baselines (% baseline period, $n = 9$), and (dark bars) D-AP5 experiments (% pre D-AP5, $n = 5$).

Methods

Electrophysiology. Thalamocortical slices³⁰ were prepared from pups (P3–8) obtained from timed-pregnant Wistar rats (P0 is defined as the day of birth) as described^{24,25}. The composition of the extracellular solution was as follows (mM): 119 NaCl, 2.5 KCl, 1.0 NaH₂PO₄, 26.2 NaHCO₃, 2.5 CaCl₂, 1.3 MgSO₄, 11 glucose, saturated with 95% O₂/5% CO₂. Most of the experiments were performed at room temperature (23–25 °C); in a few experiments the bath temperature was 28–32 °C. Whole-cell patch-clamp recordings were made from layer-IV neurons^{23–25} using patch electrodes (3–5 M Ω) filled with an intracellular solution of the following composition (mM): 135 caesium methyl sulphate, 8 NaCl, 10 HEPES, 0.5 EGTA, 4 Mg-ATP, 0.3 Na-GTP, pH 7.25, 285 mOsm. In some experiments, QX-314 (5 mM) was also included. All recordings were made in the presence of 50 μ M picrotoxin in the extracellular solution to block GABA_A-receptor-mediated currents. Caesium ions in the intracellular solution blocked GABA_B-receptor-mediated responses. In all cells, except in the LTP experiments and associated controls, 100 μ M D-AP5 was also included in the extracellular solution to block NMDAR-mediated currents. EPSCs were evoked by electrical stimulation at a frequency of 0.2 or 0.1 Hz by a bipolar stimulating electrode, placed either in the ventrobasal complex of the thalamus ($n = 32$) or in the white matter ($n = 96$). In about half of the neurons, both pathways were alternatively stimulated. White-matter pathways gave identical results to ventrobasal pathways, so data from both were pooled. Minimal stimulation was used in most experiments ($n = 82$). As previously described, these stimulation parameters selectively activate thalamocortical axons with no contamination from antidromic activation of layer IV neurons^{23–25,30}. The lack of antidromic activation was confirmed by latency analysis. For the LTP experiments and interleaved controls, we used minimal stimulation in the ventrobasal complex ($n = 25$) and EPSCs were evoked at a frequency of 0.5 or 0.33 Hz. LTP was induced by pairing stimulation (at baseline frequency) with a holding potential of 0 mV for 50 or 100 trials^{24,25}. For the interleaved D-AP5 controls, it was confirmed that the D-AP5 had washed in by checking that the NMDAR-mediated component of transmission recorded at depolarized potentials was blocked. Data was only included for EPSCs that had a smooth rising phase. All numbers of observations represent the number of pathways except when stated otherwise. Recordings were made using an Axopatch 200B; signals were filtered at 5 kHz, digitized at 10 kHz and stored on computer. EPSC amplitude, input resistance and series resistances were analysed and displayed on-line. The mean series resistance for all cells in this study was 22 ± 1 M Ω , $n = 105$ cells.

Analysis. Averaged EPSCs (average of 3 or 5) were digitally filtered at 1 kHz and fitted with one or two exponentials using SigmaPlot 3.0. Each EPSC was fit with both a single and a double exponential to determine which was the best fit. For the LTP experiments and interleaved controls, EPSC averages were constructed from all EPSCs collected during baseline (average of 50–100) and after LTP (typically an average of 300–600) or after the control manipulation. Only the data collected at room temperature were included in the kinetic analysis. For I - V analysis, holding potentials of -70 , -50 , -30 , -10 , 10 , 30 and 50 mV were used; for the dual component EPSCs, separate amplitude measurements were made at the peak and at the tail ($5-6 \times \tau_{\text{fast}}$ from the peak). Fits for the I - V data were obtained using linear regression analysis (SigmaPlot 3.0); first order for peak of dual component EPSC, third order for dual EPSC tail and kainate EPSC peak. Two rectification indices were calculated for each I - V series; a ratio of the conductance (I_{EPSC}/V) for currents at $+30$ and -70 mV, and a ratio for $+50$ and -50 mV. The mean rectification index for $+40$ versus -60 mV was estimated by averaging the mean index values for the two sets of potentials (this allowed a direct comparison with data from ref. 20). Spontaneous EPSC traces were constructed by aligning 10–20 events to the point of fastest rise by eye and averaging. Charge transfer through the AMPAR- and KAR-mediated components of the dual-component EPSCs were estimated by calculating the product of τ and the current at peak obtained from the exponential fits. All values are expressed as mean $\pm$ s.e.m. Statistical significance was assessed using two-tailed paired or unpaired Student's t -tests as appropriate ($P < 0.05$ considered significant).

Drugs. The active isomer of GYKI 53655 was used in all experiments (supplied by Eli Lilly). D-AP5, CNQX (Tocris) and picrotoxin (Sigma) were obtained commercially.

Received 6 May; accepted 15 June 1999.

- Hollmann, M. & Heinemann, S. Cloned glutamate receptors. *Annu. Rev. Neurosci.* **17**, 31–108 (1994).
- Bettler, B. & Mülle, C. Neurotransmitter receptors II. AMPA and kainate receptors. *Neuropharmacology* **34**, 123–139 (1995).
- Bleakman, D. & Lodge, D. Neuropharmacology of AMPA and kainate receptors. *Neuropharmacology* **37**, 1187–1204 (1998).
- Chittajallu, R., Braithwaite, S. P., Clarke, V. R. J. & Henley, J. M. Kainate receptors: subunits, synaptic localisation and function. *Trends Pharmacol. Sci.* **20**, 26–35 (1999).
- Chittajallu, R. *et al.* Regulation of glutamate release by presynaptic kainate receptors in the hippocampus. *Nature* **379**, 78–81 (1996).
- Clarke, V. R. J. *et al.* A hippocampal GluR5 kainate receptor regulating inhibitory synaptic transmission. *Nature* **389**, 599–603 (1997).
- Rodríguez-Moreno, A., Herreras, O. & Lerma, J. Kainate receptors presynaptically downregulate GABAergic inhibition in the rat hippocampus. *Neuron* **19**, 893–901 (1997).
- Castillo, P. E., Malenka, R. C. & Nicoll, R. A. Kainate receptors mediate a slow postsynaptic current in hippocampal CA3 neurons. *Nature* **388**, 182–186 (1997).
- Vignes, M. & Collingridge, G. L. The synaptic activation of kainate receptors. *Nature* **388**, 179–182 (1997).
- Cossart, R., Esclapez, M., Hirsch, J. C., Bernard, C. & Ben-Ari, Y. GluR5 kainate receptor activation in interneurons increases tonic inhibition of pyramidal cells. *Nature Neurosci.* **1**, 470–478 (1998).
- Frerking, M., Malenka, R. C. & Nicoll, R. A. Synaptic activation of kainate receptors on hippocampal interneurons. *Nature Neurosci.* **1**, 479–486 (1998).
- Li, H. & Rogawski, M. A. GluR5 kainate receptor-mediated synaptic transmission in rat basolateral amygdala in vitro. *Neuropharmacology* **37**, 1279–1286 (1998).
- DeVries, S. H. & Schwartz, E. A. Kainate receptors mediate synaptic transmission between cones and 'Off' bipolar cells in a mammalian retina. *Nature* **397**, 157–160 (1999).
- Li, P. *et al.* Kainate-receptor-mediated sensory synaptic transmission in mammalian spinal cord. *Nature* **397**, 161–164 (1999).
- Paternain, A. V., Morales, M. & Lerma, J. Selective antagonism of AMPA receptors unmasks kainate receptor-mediated responses in hippocampal neurons. *Neuron* **14**, 185–189 (1995).
- Wilding, T. J. & Huettner, J. E. Differential antagonism of alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid-preferring and kainate-preferring receptors by 2,3-benzodiazepines. *Mol. Pharmacol.* **47**, 582–587 (1995).
- Ruano, D., Lambolez, B., Rossier, J., Paternain, A. V. & Lerma, J. Kainate receptor subunits expressed in single cultured hippocampal neurons: molecular and functional variants by RNA editing. *Neuron* **14**, 1009–1017 (1995).
- Bowie, D. & Mayer, M. L. Inward rectification of both AMPA and kainate subtype glutamate receptors is generated by polyamine-mediated ion channel block. *Neuron* **15**, 453–462 (1995).
- Kamboj, S. K., Swanson, G. T. & Cull-Candy, S. G. Intracellular spermine confers rectification on rat calcium-permeable AMPA and kainate receptors. *J. Physiol.* **486**, 297–303 (1995).
- Tóth, K. & McBain, C. J. Afferent-specific innervation of two distinct AMPA receptor subtypes on single hippocampal interneurons. *Nature Neurosci.* **1**, 572–578 (1998).
- Washburn, M. S., Numberger, M., Zhang, S. & Dingledine, R. Differential dependence on GluR2 expression of three characteristic features of AMPA receptors. *J. Neurosci.* **17**, 9393–9406 (1997).
- Fox, K. The critical period for long-term potentiation in primary sensory cortex. *Neuron* **15**, 485–488 (1995).
- Craig, M. C. & Malenka, R. C. A critical period for long-term potentiation at thalamocortical synapses. *Nature* **375**, 325–328 (1995).
- Isaac, J. T. R., Craig, M. C., Nicoll, R. A. & Malenka, R. C. Silent synapses during development of thalamocortical inputs. *Neuron* **18**, 269–280 (1997).
- Feldman, D. E., Nicoll, R. A., Malenka, R. C. & Isaac, J. T. R. Long-term depression at thalamocortical synapses in developing rat somatosensory cortex. *Neuron* **21**, 347–357 (1998).
- Kullmann, D. M. & Asztély, F. Extrasynaptic glutamate spillover in the hippocampus: evidence and implications. *Trends Neurosci.* **21**, 8–14 (1998).
- Paternain, A. V., Rodríguez-Moreno, A., Villarroel, A. & Lerma, J. Activation and desensitisation properties of native and recombinant kainate receptors. *Neuropharmacology* **37**, 1249–1260 (1998).

- Patneau, D. K. & Mayer, M. L. Structure–activity relationships for amino acid transmitter candidates acting at N -methyl-D-aspartate and quisqualate receptors. *J. Neurosci.* **10**, 2385–2399 (1990).
- Rozov, A., Zilberter, Y., Wollmuth, L. P. & Burnashev, N. Facilitation of currents through rat Ca^{2+} -permeable AMPA receptor channels by activity-dependent relief from polyamine block. *J. Physiol.* **511**, 361–377 (1998).
- Agmon, A. & Connors, B. W. Thalamocortical responses of mouse somatosensory (barrel) cortex in vitro. *Neuroscience* **41**, 365–379 (1991).

Acknowledgements. We thank G. Collingridge, V. Clarke, R. Chittajallu, S. Braithwaite and J. Crabtree for advice and comments; W. W. Anderson for providing the data acquisition software; and D. Lodge (Eli Lilly and Co.) for supplying the GYKI 53655. This work was supported by the Wellcome Trust.

Correspondence and requests for materials should be addressed to J.L.

Oligopeptide-repeat expansions modulate 'protein-only' inheritance in yeast

Jia-Jia Liu & Susan Lindquist

Howard Hughes Medical Institute, Department of Molecular Genetics and Cell Biology, University of Chicago, 5841 S. Maryland Avenue MC1028, Chicago, Illinois 60637, USA

The yeast $[\text{PSI}^+]$ element represents a new type of genetic inheritance, in which changes in phenotype are transmitted by a 'protein only' mechanism^{1–3} reminiscent of the 'protein-only' transmission of mammalian prion diseases^{1,4}. The underlying molecular mechanisms for both are poorly understood and it is not clear how similar they might be. Sup35, the $[\text{PSI}^+]$ protein determinant, and PrP, the mammalian prion determinant, have different functions, different cellular locations and no sequence similarity; however, each contains five imperfect oligopeptide repeats—PQGGYQQYN in Sup35 and PHGGGWGQ in PrP^{Sc}. Repeat expansions in PrP produce spontaneous prion diseases^{7,8}. Here we show that replacing the wild-type SUP35 gene with a repeat-expansion mutation induces new $[\text{PSI}^+]$ elements, the first mutation of its type among these newly described elements of inheritance. *In vitro*, fully denatured repeat-expansion peptides can adopt conformations rich in β -sheets and form higher-order structures much more rapidly than wild-type peptides. Our results provide insight into the nature of the conformational changes underlying protein-based mechanisms of inheritance and suggest a link between this process and those producing neurodegenerative prion diseases in mammals.

The Sup35 protein of *Saccharomyces cerevisiae* is a subunit (eRF3) of the eukaryotic release factor^{9,10}. In $[\text{PSI}^+]$ cells, most of the Sup35 protein is sequestered into higher-order protein complexes and is unavailable to function in translation termination^{2,3}. Once formed, these complexes are self-perpetuating and are passed from mother cells to daughters. As a result, $[\text{PSI}^+]$ causes a heritable change in the fidelity of protein synthesis: ribosomes have an increased tendency to read through stop codons, and nonsense mutations are suppressed. Certain point mutations and deletions in the amino-terminal prion-determining domain of Sup35 (NPD) are defective in the production of higher-order Sup35 complexes and in the propagation of $[\text{PSI}^+]$ ^{11–14}. To investigate the importance of the oligopeptide repeats in this region, we created one variant with the last four repeats deleted (RΔ2–5) and another with two additional copies of the second repeat (R2E2). The wild-type gene was replaced by the genes encoding these variants, in its normal chromosomal context, by homologous recombination in isogenic $[\text{PSI}^+]$ and $[\text{psi}^-]$ cells. Western blotting confirmed that all three proteins accumulated to the same extent.

First, we assayed the spontaneous appearance of new $[\text{PSI}^+]$ elements in $[\text{psi}^-]$ strains. The *ade1–14* allele, which contains a

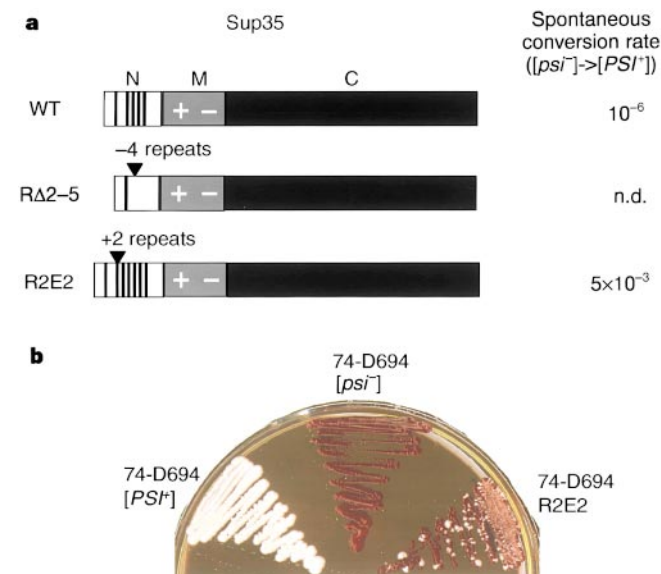


Figure 1 Effect of the R2E2 and R Δ 2-5 mutations on spontaneous conversion from $[psi^-]$ to $[PSI^+]$. **a**, Diagrams of wild-type (WT) and mutant Sup35 and table of conversion rates determined by plating cells on SD-Ade medium. Vertical lines in the N-terminal region of Sup35 (N) represent oligopeptide repeats. The highly charged nature of the middle region (M) is indicated. The C-terminal region (C) functions in translation termination. n.d., no guanidine HCl-curable Ade⁺ cells were detected in $>10^8$ cells plated. **b**, White $[PSI^+]$ colonies appear at a high rate when $[psi^-]$ cells with an R2E2/SUP35 replacement are streaked on YPD.

nonsense mutation in the auxotrophic marker *ADE1*, provides a particularly convenient method for monitoring the appearance of $[PSI^+]$. The suppression of *ade1-14* allows cells to grow on adenine-deficient synthetic medium (SD-Ade) and prevents the accumulation of metabolic by-products that cause colonies to appear red on rich medium (YPD)^{12,15}. Wild-type cells of strain 74-D694 rarely produced colonies on SD-Ade medium (frequency was approximately one per 10^6 cells plated) and rarely produced white colonies on rich medium (Fig. 1). R Δ 2-5 cells produced such colonies at an even lower frequency (less than 1 in 10^8 cells); none of these was due to the appearance of $[PSI^+]$, because they could not be cured by growth on medium containing 5 mM guanidine hydrochloride¹⁶. In striking contrast, R2E2 cells frequently produced colonies on SD-Ade medium ($\sim 5,000$ times more frequently than wild-type cells), and similarly increased the spontaneous appearance of white colonies on rich medium (Fig. 1). Growth on medium containing guanidine-HCl cured both the ability of the cells to grow on SD-Ade medium and the white colony colour on rich medium, confirming that the phenotypes were due to the appearance of genuinely new $[PSI^+]$ elements.

Next we examined the effects of the repeat variants when they replaced the wild-type *SUP35* gene in cells that were already $[PSI^+]$. Cells with the R2E2 replacement remained $[PSI^+]$. All cells carrying the R Δ 2-5 replacement became $[psi^-]$ (data not shown). Thus, the oligopeptide repeats are crucial for the maintenance of $[PSI^+]$.

Two defining features of the $[PSI^+]$ phenomenon are that transient overexpression of fragments carrying the N-terminal region of Sup35 is sufficient to induce heritable new $[PSI^+]$ elements^{11,12,17}, and the appearance of $[PSI^+]$ is associated with the formation of higher-order complexes of Sup35^{2,3,13,14}. We examined whether transient expression of N-terminal fragments containing the repeat variants could induce new $[PSI^+]$ elements in wild-type cells, and also whether this would correlate with the appearance of new protein complexes. To do this, we used a fusion between NM, the N-terminal and middle domains of Sup35 (amino acids 1–253),

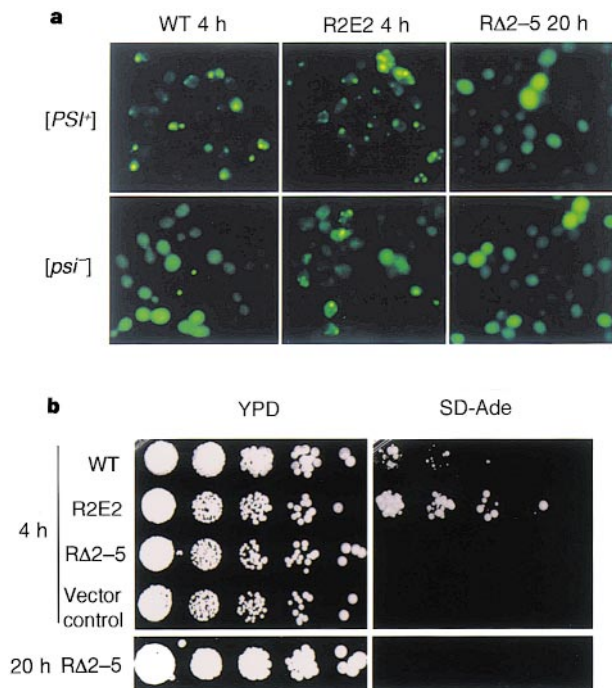


Figure 2 Visualization of $[PSI^+]$ conversions. **a**, Fluorescence of NM-GFP fusion proteins in $[PSI^+]$ and $[psi^-]$ cells after 4 h (wild-type (WT) NM-GFP and R2E2 NM-GFP) and 20 h (R Δ 2-5 NM-GFP) of induction. **b**, Induction of heritable $[PSI^+]$ factors. $[psi^-]$ cells analysed in **a** were serially diluted in 5-fold increments and spotted onto YPD medium, to measure total cells in the culture, and SD-Ade medium, to measure $[PSI^+]$ cells. Control construct, GFP alone.

and green fluorescent protein (NM-GFP). In $[PSI^+]$ cells, newly synthesized NM-GFP rapidly coalesces into intense fluorescent foci as it joins pre-existing complexes of Sup35². In $[psi^-]$ cells, NM-GFP is initially soluble, but prolonged expression produces intense fluorescent foci coincident with the formation of new $[PSI^+]$ elements².

Plasmids encoding inducible wild-type and repeat-variant NM-GFPs were transformed into isogenic $[PSI^+]$ and $[psi^-]$ cells with a wild-type *SUP35* gene in the chromosome. R Δ 2-5 NM-GFP produced no intense fluorescent foci in either $[PSI^+]$ or $[psi^-]$ cells even after 20 h of induction. Both R2E2 NM-GFP and wild-type NM-GFP produced intense foci in virtually all $[PSI^+]$ cells after only 4 h of induction (Fig. 2a). New fluorescent foci appeared in $[psi^-]$ cells at a higher frequency with R2E2 NM-GFP than with wild-type NM-GFP at both 4 and 20 h (Fig. 2a, and data not shown). To monitor the appearance of new $[PSI^+]$ elements in the $[psi^-]$ cells, they were plated onto media selective for $[PSI^+]$ but not for the plasmids. No $[PSI^+]$ colonies developed from cultures that had expressed R Δ 2-5 NM-GFP for 20 h (Fig. 2b). Cells expressing R2E2 NM-GFP for 4 and 20 h produced ~ 5 -fold as many $[PSI^+]$ colonies as cells expressing wild-type NM-GFP (Fig. 2b, and data not shown).

The results suggest that the oligopeptide repeat expansion increases the spontaneous appearance of new $[PSI^+]$ elements by facilitating the initial conformational conversion of Sup35 protein and directly influencing the conversion of newly made protein. But it remains possible that R2E2 NM increases the appearance of $[PSI^+]$ elements indirectly by affecting the physiology of the cell in a manner that in turn influences the independent folding of other Sup35 molecules. (Indeed, this is a caveat for all previous investigations examining the conversion of $[psi^-]$ cells to $[PSI^+]$.) To determine whether proteins expressed from the plasmids interact directly with endogenous Sup35, we used a GFP antibody and immunomagnetic beads to extract proteins associated with the GFP

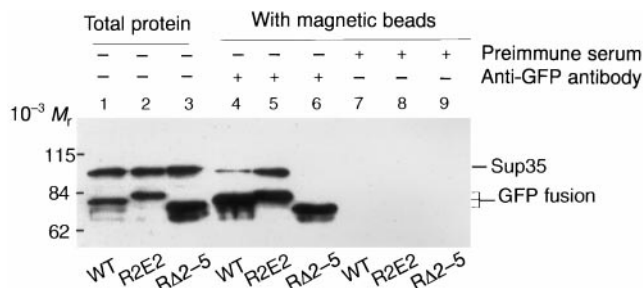


Figure 3 Interactions between endogenous full-length Sup35 and GFP-fusion proteins. [*psi*⁻] cells were induced to express NM-GFP proteins with CuSO₄ for 4 h. GFP-fusion proteins were extracted from yeast cell lysates by purification with immunomagnetic beads by using an anti-GFP antibody. Protein complexes were analysed by SDS-PAGE and immunoblotting with anti-Sup35 1B antibody. Lanes 1–3, total proteins in the lysate; lanes 4–6, proteins bound to magnetic beads in the presence of anti-GFP antibodies; lanes 7–9, proteins bound in the presence of preimmune serum. The positions of relative molecular mass (*M_r*) markers and of wild-type (WT) Sup35 and NM-GFP fusions are indicated.

fusions from cell lysates. When wild-type NM-GFP was induced in [*psi*⁻] cells for 4 h, some wild-type Sup35 was recovered on the beads (Fig. 3). With R2E2 NM-GFP, a larger fraction was recovered, and with RA2-5 NM-GFP, none. The increased appearance of new [*PSI*⁺] elements in cells that express R2E2 NM-GFP is probably due to the protein's enhanced ability to self-assemble into higher-order complexes that can recruit and convert new Sup35 protein to the [*PSI*⁺] state.

Next, we tested the effects of the repeat variants on the structural transitions of NM *in vitro*. When purified recombinant NM is denatured and diluted into aqueous buffers, it slowly changes from a random coil to a structure rich in β -sheets and forms fibres that bind Congo red, giving the spectral shift characteristic of amyloid proteins¹³. When deposited at high concentrations, these Congo-red-stained fibres also show apple-green birefringence (A. Cashikar and S.L., unpublished observations). The ability of preformed β -sheet-rich structures to promote the rapid conversion of soluble NM to the same state provides a simple molecular model for the propagation of [*PSI*⁺] *in vivo*^{13,14,18}. To determine whether the repeat variants alter the intrinsic capacity of the protein to fold into this form, we purified wild-type NM proteins and the repeat variants in the fully denatured state and diluted them into non-denaturing buffer. Structural changes were monitored by the binding of Congo red (Fig. 4) and confirmed by circular dichroism and electron microscopy (data not shown). The three proteins converted at profoundly different rates: the repeat expansion variant converted much more rapidly than the wild type, whereas the deletion variant converted much more slowly. The differences were obtained reproducibly in both rotated and unrotated reactions, although the rate of conversion in unrotated reactions was slower overall (data not shown). Thus, the mutations alter the intrinsic propensity of the protein, starting from the unfolded state to progress to a β -sheet-rich, more highly ordered structure.

Like the mammalian prion, [*PSI*⁺] is believed to propagate by a 'protein only' mechanism. We find that an expansion of the oligopeptide repeats induces the spontaneous appearance of [*PSI*⁺] and that deletion of the repeats eliminates it. Notably, certain mammalian prion diseases are linked to expansions of the oligopeptide repeats in PrP^{7,8}, and variants with reduced numbers of repeats are less susceptible to prion disease¹⁹ (C. Weissmann, personal communication). The extraordinary similarity of the mutations that affect, on the one hand, the spontaneous appearance of a heritable proteinaceous element in yeast and, on the other hand, a heritable protein misfolding disease in mammals suggests a congruity in their underlying molecular mechanisms. In both fields, a major unanswered question is whether the acquisition of

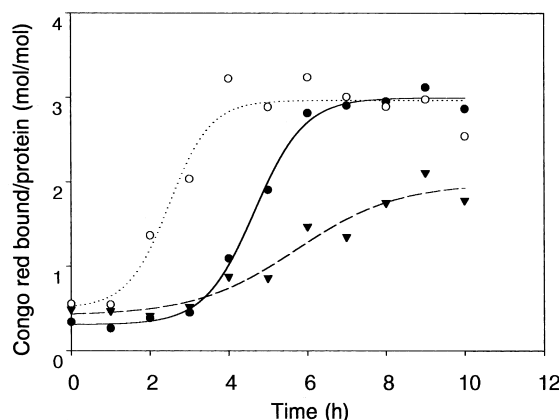


Figure 4 Kinetics of fibre formation of NM repeat mutants *in vitro*. Recombination proteins, wild-type (filled circles), R2E2 (open circles) or RA2-5 (triangles), purified from *Escherichia coli*²⁶ were diluted into aqueous buffer and the progression of fibre formation was monitored by the spectral shift produced on Congo-red binding²⁹. Sigmoid regression analysis was done with Sigmaplot for Windows Version 4.01 (SPSS). Correlation coefficients: wild type, 0.9977; R2E2, 0.9749; RA2-5, 0.9720.

the prion state occurs through the structural rearrangement of native protein or through the diversion of transiently unfolded (or nascent) protein to an alternative folding pathway. Our experiments *in vitro* started with denatured protein. They therefore indicate that the rate at which higher-order, β -sheet-rich structure is acquired after the protein has reached the unfolded state can be a rate-limiting factor in the creation of new [*PSI*⁺] elements. PrP mutations are thought to cause disease by decreasing the stability of native PrP; there is strong evidence that this is an important general mechanism by which mutations can lead to protein-misfolding diseases^{20,21}. But recent work on folded PrP indicates that at least some mutants have the same thermodynamic stability as wild-type PrP^{22,23}. We suggest that some disease mutations enhance amyloidogenic processes by altering the forward folding and assembly pathway after energy barriers to unfolding have been overcome. □

Methods

Plasmid constructions. NM-GFP fusion constructs of R2E2 and RA2-5 were created by site-directed mutagenesis, using oligonucleotide primers and QuickChange Site-Directed Mutagenesis Kit (Stratagene). Primers for R2E2 were 5'-CAAGGTGGCTATCAACAGTACAATCCCCAAGGTGGCTATCAACAG-3' and 5'-CTGTTGATAGCCACCTTGGGGATTGTACTGTTGATAGCCACCTTG-3'. Primers for RA2-5 were 5'-ATTCTGGGTACCAACCACAAGGTGGCCGTG-3' and 5'-CACGGCCACCTTGTGGTTGGTACCCAGAAT-3'. The template plasmid for mutagenesis was p316CUP1-3SGFP^{SG}. To create p316CUP1-3SGFP^{SG}, the CUP1 promoter sequence from pCLUC (gift from D. Thiele)²⁴ was amplified by PCR and inserted into pRS316²⁵ at *EcoRI*/*Bam*HI sites using PCR primers 5'-CGGAATTCCTTACCGACATTTGGGCGCT-3' and 5'-CGGGATCCTGATTGATTGATTGTACAGTTTG-3'. Wild-type NM was amplified by PCR and inserted into *Bam*HI/*Sac*I sites (primers: 5'-CGCGGATCCATGTGCGATTCAAACC-3' and 5'-CGCGAGTCACTAGCTAGTT-3'). GFP^{SG} was amplified by PCR using pJK19-1 (gift from P. Silver) as a template and inserted into *Sac*I/*Sac*I sites (primers: 5'-GATGAGCTCATGGCTAGCAAAGGAG-3' and 5'-TCCCCGCGGTCATCCTTTGTATAGTTTCATCCAT-3').

A SUP35 integrative vector was constructed with SUP35 genomic sequences (5' flanking region 1360 nt; 3' flanking region 800 nt) inserted at *Xho*I/*Eco*RI and *Bam*HI/*Sac*II sites in pRS306²⁵, respectively, and was named pJLI-SUP35 (PCR primers for 5' flanking region: 5'-CAGCAACTCGAGAAGATATCCATCAT-3' and 5'-CGGAATTCTGTTGCTAGTGGGCAGAT-3'; primers for 3' flanking region: 5'-CGGGATCCATTTCTTGCAACATAAGTAAATGCAAAC-3' and 5'-TCCCCGCGGTGAAAGAGTCACTGAGACGACGACT-3'). The DNA sequence encoding the Sup35 carboxy terminus (amino acids 124–685)

was amplified by PCR and inserted into pJLI-SUP35 at *EcoRI/BamHI* sites (primers: 5'-CGGAATTCATGCTCTTGAACGACTTCAAAAGC-3' and 5'-CGCGGATCCTTACTCGGCAATTTTAAC-3'). The R2E2 and RΔ2-5 NPΔ sequences were then amplified and inserted into the plasmid at the *EcoRI* site, respectively (PCR primers: 5'-CGGAATTCATGCTCGGATTCAAACCAAGG-3' and 5'-CGGAATTCACCTTGAGACTGTGGTTGG-3'). The two constructs were then named pJLI-SUP35 R2E2 and pJLI-SUP35 RΔ2-5.

The non-His-tagged bacterial expression constructs of R2E2 and RΔ2-5 were constructed in the same way as the wild-type NM expression construct²⁶. All constructs were confirmed by dideoxy nucleotide triphosphate sequencing. **Gene integration and replacement.** Integrative constructs were digested with *XbaI* and transformed into 74-D694 [*PSI*⁺] and [*psi*⁻] strains. Transformants were selected on uracil-deficient (SD-Ura) medium and confirmed by genomic PCR. Recombinant excision events were selected on medium containing 5-fluoro-orotic acid²⁷. Strains in which wild-type SUP35 was replaced with the R2E2 and RΔ2-5 mutations were screened by PCR and confirmed by western blotting.

Spontaneous conversion from [*psi*⁻] to [*PSI*⁺]. To measure spontaneous conversion from [*psi*⁻] to [*PSI*⁺], red colonies carrying wild-type SUP35, R2E2 or RΔ2-5 were inoculated into liquid YPD medium and grown overnight at 30 °C. Cells were adjusted to equal densities, serially diluted and plated on to SD-Ade medium (5 × 10⁻²–10⁻⁸ cells per plate, depending on the strain). To confirm the presence of [*PSI*⁺], colonies were patched onto YPD medium containing 5 mM guanidine-HCl, as described^{16,28}.

Analysis of NM-GFP cells. An exponential-phase isogenic pair of 74-D694 [*PSI*⁺] and [*psi*⁻] cells was induced to express the GFP-fusion proteins with 50 μM CuSO₄ at 30 °C. For fluorescence microscopy, cells were fixed with 1% formaldehyde after 4 and 20 h. For analysis of [*PSI*⁺] induction, [*psi*⁻] cells overexpressing GFP fusion proteins were serially diluted and spotted onto YPD and SD-Ade media after 4 and 20 h. The insolubility of Sup35 as the protein converts to the [*PSI*⁺] state obviates standard immunoprecipitation procedures to identify protein complexes. For immunomagnetic purification of NM-GFP::Sup35 complexes, [*psi*⁻] cells were suspended in protoplasting buffer (1 M sorbitol, 0.1 M EDTA, 50 mM dithiothreitol and 0.2 mg ml⁻¹ zymolyase) after 4 h of induction and incubated at 37 °C for 1 h. Protoplasts were lysed in 50 mM Tris-HCl, pH 7.5, containing 5 mM MgCl₂, 10 mM KCl, 0.1 mM EDTA, 2 mM PMSF, 100 μg ml⁻¹ ribonuclease A, 10 μg ml⁻¹ leupeptin, 2 μg ml⁻¹ pepstatin and 1 mM benzamidin. All samples were adjusted to 2 mg ml⁻¹ total protein (Bio-Rad Bradford assay). Aliquots of each sample were then incubated with anti-GFP antibody (rabbit polyclonal; Clontech) or preimmune rabbit serum at 4 °C for 1 h. The antibody–protein complexes were isolated with magnetic beads (Dynabeads M-280 sheep anti-rabbit IgG; DYNAL) and washed. Bound proteins were eluted with SDS-sample buffer, analysed by immunoblotting with anti-Sup35 antibody 1B, which recognizes the middle region of Sup35², and detected by chemiluminescence (ECL western blotting reagents; Amersham).

Congo-red binding. Congo-red binding was performed as described²⁹. Denatured bacterial recombinant proteins were precipitated with 67% methanol and then redissolved in Congo red binding buffer (5 mM potassium phosphate, pH 7.4, 150 mM NaCl) to a final concentration of 10 μM. Proteins were then subjected to continuous slow rotation at a speed of 2.5 r.p.m. At indicated times, triplicate aliquots of each reaction were diluted to 2 μM protein and incubated with 10 μM Congo red. Congo-red binding was calculated as previously described²⁹ and plotted as the mean of triplicate determinations.

Received 23 March; accepted 9 June 1999.

- Wickner, R. B. [URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. *Science* **264**, 566–569 (1994).
- Patino, M. M., Liu, J. J., Glover, J. R. & Lindquist, S. Support for the prion hypothesis for inheritance of a phenotypic trait in yeast. *Science* **273**, 622–626 (1996).
- Paushkin, S. V., Kushnirov, V. V., Smirnov, V. N. & Ter-Avanesyan, M. D. Propagation of the yeast prion-like [*psi*⁻] determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor. *EMBO J.* **15**, 3127–3134 (1996).
- Lindquist, S. Mad cows meet psi-chotic yeast: the expansion of the prion hypothesis. *Cell* **89**, 495–498 (1997).
- Kretschmar, H. A. *et al.* Molecular cloning of a human prion protein cDNA. *DNA* **5**, 315–324 (1986).
- Kushnirov, V. V. *et al.* Nucleotide sequence of the SUP2 (SUP35) gene of *Saccharomyces cerevisiae*. *Gene* **66**, 45–54 (1988).
- Prusiner, S. B. & Scott, M. R. Genetics of prions. *Annu. Rev. Genet.* **31**, 139–175 (1997).
- Chiesa, R., Piccardo, P., Ghetti, B. & Harris, D. A. Neurological illness in transgenic mice expressing a prion protein with an insertional mutation. *Neuron* **21**, 1339–1351 (1998).

- Zhouravleva, G. *et al.* Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *EMBO J.* **14**, 4065–4072 (1995).
- Stansfield, I. *et al.* The products of the SUP45 (eRF1) and SUP35 genes interact to mediate translation termination in *Saccharomyces cerevisiae*. *EMBO J.* **14**, 4365–4373 (1995).
- Ter-Avanesyan, M. D. *et al.* Deletion analysis of the SUP35 gene of the yeast *Saccharomyces cerevisiae* reveals two non-overlapping functional regions in the encoded protein. *Mol. Microbiol.* **7**, 683–692 (1993).
- Derkatch, I. L., Chernoff, Y. O., Kushnirov, V. V., Inge-Vechtomov, S. G. & Liebman, S. W. Genesis and variability of [*PSI*] prion factors in *Saccharomyces cerevisiae*. *Genetics* **144**, 1375–1386 (1996).
- Glover, J. R. *et al.* Self-seeded fibers formed by Sup35, the protein determinant of [*PSI*⁺], a heritable prion-like factor of *S. cerevisiae*. *Cell* **89**, 811–819 (1997).
- DePace, A. H., Santos, A., Hillner, P. & Weissman, J. S. A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. *Cell* **93**, 1241–1252 (1998).
- Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtomov, S. G. & Liebman, S. W. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [*psi*⁺]. *Science* **268**, 880–884 (1995).
- Tuite, M. F., Mundy, C. R. & Cox, B. S. Agents that cause a high frequency of genetic change from [*psi*⁺] to [*psi*⁻] in *Saccharomyces cerevisiae*. *Genetics* **98**, 691–711 (1981).
- Chernoff, Y. O., Derkach, I. L. & Inge-Vechtomov, S. G. Multicopy SUP35 gene induces de-novo appearance of *psi*-like factors in the yeast *Saccharomyces cerevisiae*. *Curr. Genet.* **24**, 268–270 (1993).
- Paushkin, S. V., Kushnirov, V. V., Smirnov, V. N. & Ter-Avanesyan, M. D. *In vitro* propagation of the prion-like state of yeast Sup35 protein. *Science* **277**, 381–383 (1997).
- Goldmann, W., Chong, A., Foster, J., Hope, J. & Hunter, N. The shortest known prion protein gene allele occurs in goats, has only three octapeptide repeats and is non-pathogenic. *J. Gen. Virol.* **79**, 3173–3176 (1998).
- McCutchen, S. L., Lai, Z., Mirov, G. J., Kelly, J. W. & Colon, W. Comparison of lethal and nonlethal transhyretin variants and their relationship to amyloid disease. *Biochemistry* **34**, 13527–13536 (1995).
- Booth, D. R. *et al.* Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**, 787–793 (1997).
- Swietnicki, W., Petersen, R. B., Gambetti, P. & Surewicz, W. K. Familial mutations and the thermodynamic stability of the recombinant human prion protein. *J. Biol. Chem.* **273**, 31048–31052 (1998).
- Liemann, S. & Glockshuber, R. Influence of amino acid substitutions related to inherited human prion diseases of the thermodynamical stability of the cellular prion protein. *Biochemistry* **38**, 3258–3267 (1999).
- Thiele, D. J. ACE1 regulates expression of the *Saccharomyces cerevisiae* metallothionein gene. *Mol. Cell. Biol.* **8**, 2745–2752 (1988).
- Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).
- Serio, T. R., Cashikar, A. G., Moslehi, J. J., Kowal, A. S. & Lindquist, S. D. The yeast prion [*PSI*⁺] and its determinant, Sup35p. *Methods Enzymol.* (in the press).
- Ausubel, F. M. *et al.* *Current Protocols in Molecular Biology* (Greene Publishing Associates/Wiley Interscience, New York, 1991).
- Derkatch, I. L., Bradley, M. E., Zhou, P., Chernoff, Y. O. & Liebman, S. W. Genetic and environmental factors affecting the *de novo* appearance of the [*PSI*⁺] prion in *Saccharomyces cerevisiae*. *Genetics* **147**, 507–519 (1997).
- Klunk, W. E., Pettigrew, J. W. & Abraham, D. J. Two simple methods for quantifying low-affinity dye-substrating binding. *J. Histochem. Cytochem.* **37**, 1293–1297 (1989).

Acknowledgements. We thank A. Kowal for technical assistance on electron microscopy, A. Cashikar for assistance with the circular dichroism study, and H. True, T. Serio, S. Uptain, T. Scheibel, L. Li and J. Ma for comments on the manuscript. This work was supported by a grant from the National Institutes of Health and the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to S.L. (e-mail: s-lindquist@uchicago.edu).

Skeletal muscle hypertrophy is mediated by a Ca²⁺-dependent calcineurin signalling pathway

Christopher Semsarian*, Ming-Jie Wu*, Yue-Kun Ju†, Tadeusz Marciniak*, Thomas Yeoh*, David G. Allen†, Richard P. Harvey*‡ & Robert M. Graham*‡

* Victor Chang Cardiac Research Institute, St Vincent's Hospital, New South Wales, 2010, Australia

† Department of Physiology, University of Sydney, New South Wales, 2006, Australia

‡ School of Biochemistry and Molecular Genetics, University of New South Wales, New South Wales 2033, Australia

Skeletal muscle hypertrophy and regeneration are important adaptive responses to both physical activity and pathological stimuli¹. Failure to maintain these processes underlies the loss of skeletal muscle mass and strength that occurs with ageing and in myopathies². Here we show that stable expression of a gene encoding insulin-like growth factor 1 (IGF-1) in C2C12 skeletal muscle cells, or treatment of these cells with recombinant IGF-1 or with insulin and dexamethasone, results in hypertrophy of

differentiated myotubes and a switch to glycolytic metabolism. Treatment with IGF-1 or insulin and dexamethasone mobilizes intracellular calcium, activates the Ca^{2+} /calmodulin-dependent phosphatase calcineurin, and induces the nuclear translocation of the transcription factor NF-ATc1. Hypertrophy is suppressed by the calcineurin inhibitors cyclosporin A or FK506, but not by inhibitors of the MAP-kinase or phosphatidylinositol-3-OH kinase pathways. Injecting rat latissimus dorsi muscle with a plasmid encoding IGF-1 also activates calcineurin, mobilizes satellite cells and causes a switch to glycolytic metabolism. We propose that growth-factor-induced skeletal-muscle hypertrophy and changes in myofibre phenotype are mediated by calcium mobilization and are critically regulated by the calcineurin/NF-ATc1 signalling pathway.

IGF-1 mediates the growth of muscle and other tissues, and in cultured myoblasts it is a potent proliferative and differentiating agent³. To test whether IGF-1 can also produce myotube hypertrophy, the C2C12 myoblast cell line, derived from adult mouse satellite cells, was stably transfected with a plasmid encoding IGF-1 (ref. 4). The derived clonal cell lines secreted IGF-1 into the culture medium and, after differentiation in low serum medium, exhibited myotube hypertrophy (Fig. 1a)—that is, increased myotube size (mean area (Fig. 1b), and width (data not shown)) and increased protein synthesis without increased DNA replication (Fig. 1c). No hypertrophy was observed when C2C12 cells were treated with another growth factor, neuregulin (50–100 nM), which like IGF-1 and insulin, acts through tyrosine kinase receptors. We also evaluated the metabolic status of IGF-1 myotubes. Selective enlargement of fast-twitch glycolytic myofibres occurs with high-resistance isotonic exercise in human⁵ and in some transgenic mouse models of skeletal muscle hypertrophy⁶, and this fibre-type is selectively lost with ageing and in animal models of muscular dystrophy⁷. Both the activity of glycolytic enzymes and the end-product of glycolytic metabolism, lactate, were significantly increased in IGF-1 myotubes⁴ (Fig. 1b). As IGF-1-transfected C2C12 cells were clonal, our findings indicate that IGF-1 not only induces muscle hypertrophy, but also mediates an adaptive developmental switch in myofibre phenotype.

It has been suggested that the Ca^{2+} /calmodulin-dependent phosphatase, calcineurin, which mediates activation of B and T lymphocytes, is involved in the induction of cardiac hypertrophy⁷. However, this issue remains controversial, with cardiac hypertrophy in some animal models being resistant to calcineurin inhibition^{8,9}. We found that steady-state levels of the 60K (relative molecular mass, M_r , of 60,000) catalytic subunit (CnA) and 19K regulatory subunit (CnB) of calcineurin were unaltered in IGF-1 myotubes (Fig. 2a). However, compared to control C2C12 cells transfected with vector only, calcineurin phosphatase (CnPP) activity was increased in IGF-1 myoblasts and remained high with differentiation to myotubes (Fig. 2a). In these studies, CnPP activity was determined in the presence of added calmodulin (100 nM) and calcium (0.1 mM) in a standard buffer that provides a free Ca^{2+} concentration of ~ 140 nM. Detailed evaluation of CnPP kinetics in C2C12 cell extracts indicated that calmodulin is not rate limiting, CnPP activity is directly related to free Ca^{2+} concentration, and the standard assay conditions accurately measure intracellular enzyme activity in both IGF-1-treated and control cells. Therefore, in the absence of changes in CnA or CnB expression, increased CnPP activity is consistent with IGF-1-induced calcium mobilization. Such an effect has not been reported previously in skeletal muscle to our knowledge. However, IGF-1 can regulate L-type Ca^{2+} -channel expression^{10,11} and Ca^{2+} current¹¹ in rodent skeletal muscle cells, and increase intracellular calcium concentration ($[\text{Ca}^{2+}]_i$) in BALB/c 3T3 fibroblasts¹². We determined steady-state $[\text{Ca}^{2+}]_i$ directly using the lipophilic Ca^{2+} -sensitive indicator indo-1-AM and found that it was significantly higher (166 ± 16 nM, $P < 0.0005$) in IGF-1-transfected myotubes (16 h post-differentia-

tion), than in native C2C12 myotubes (94 ± 9 nM) or in control myotubes transfected with vector alone (79 ± 8 nM). Culturing cells in a Ca^{2+} -free extracellular solution, or in the presence of the L-type Ca^{2+} -channel blocker nifedipine (10 μM), indicated that the increase in $[\text{Ca}^{2+}]_i$ in IGF-1-transfected myotubes was due to calcium influx, but not through L-type Ca^{2+} channels. Coupling of calcium mobilization to CnPP activation was demonstrated by pretreating cells with the intracellular calcium chelator BAPTA-AM (1,2-bis(o-aminophenoxy)ethane- N,N,N',N' -tetraacetic acid tetra(acetoxymethyl) ester). This completely inhibited CnPP activity (Fig. 2b).

To investigate whether calcineurin activation was causally related

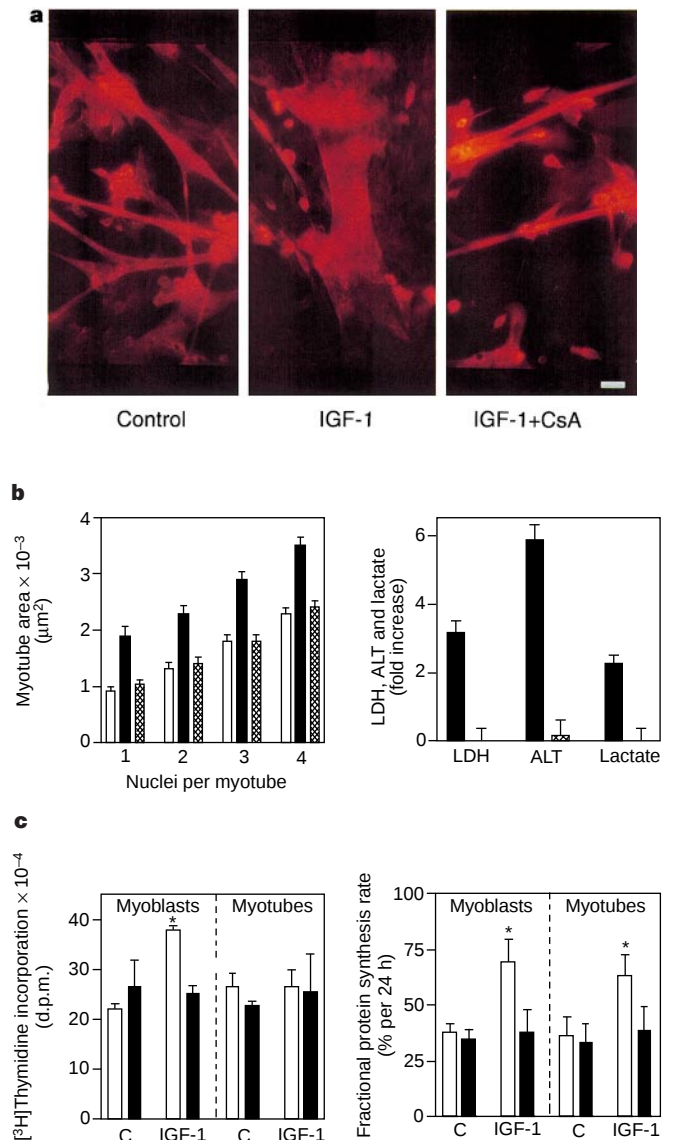


Figure 1 Induction and calcineurin-dependence of IGF-1-mediated hypertrophy in C2C12 cells. **a**, Phalloidin-stained control (vector alone), IGF-1 transfected (IGF-1) and IGF-1-plus-cyclosporin A (CsA, 1 μM , added when myoblasts were plated)-treated myotubes, five days post-differentiation (scale bar, 20 μm). **b**, Mean area of control (white bars), IGF-1 (black bars) and IGF-1-plus-CsA-treated (hatched bars) five-day post-differentiation myotubes, and glycolytic enzyme activities (LDH, lactate dehydrogenase; ALT, alanine aminotransferase) and lactate levels for IGF-1 (black bars) and IGF-1-plus-CsA (hatched bars)-treated myotubes. Myotube area and width (data not shown), as well as glycolytic enzyme and lactate levels, were greater ($P < 0.001$) than the corresponding control and IGF-1-plus-CsA values. **c**, $[\text{H}^3]$ thymidine incorporation and fractional protein synthesis rate in control (C), IGF-1 myoblasts and IGF-1 myotubes in the absence (white bars) or presence (black bars) of CsA (1 μM) (* $P < 0.01$ vs CsA-treated).

to IGF-1-induced myotube hypertrophy, cells were treated with cyclosporin A to inhibit calcineurin. As shown in Fig. 1, cyclosporin A (at 1 μM) prevented all measures of myotube hypertrophy, including the IGF-1-induced switch to glycolytic metabolism. The structurally distinct calcineurin inhibitor FK506 (at 100 ng ml^{-1}) had similar effects (data not shown). Cyclosporin A (0.5–1.0 μM) had no effect on differentiation of control or IGF-1-transfected cells, or on proliferation ($[\text{H}^3]$ thymidine uptake) of control myoblasts (<1% inhibition at 3 μM). However, cyclosporin A inhibited IGF-1-induced proliferation of transfected myoblasts by 91% at 1 μM and 95% at 3 μM . Inhibition of IGF-1-induced proliferation by cyclosporin A appears to be unrelated to its effects on myotube hypertrophy, as inhibition of the MAP-kinase pathway with PD98059 (a MEK-1 inhibitor) also suppressed IGF-1-induced proliferation (by 93% at 100 μM), but did not prevent myotube hypertrophy or the switch to glycolytic metabolism (data not shown). Inhibitors of the phosphatidylinositol-3-OH-kinase pathway (LY294002 and rapamycin), at doses (10 μM and 1 $\mu\text{g ml}^{-1}$, respectively, added during the myoblast phase) that completely prevent differentiation of L6 myoblasts^{13,14}, did not affect C2C12 cell differentiation or IGF-1-induced hypertrophic or proliferative responses (data not shown).

To confirm that involvement of the calcineurin pathway is not unique to the IGF-1 model, C2C12 cells were treated with insulin (I,

1–1,000 nM) and dexamethasone (D, 2.5 μM) from the time cells were switched to low-serum differentiation medium. Dexamethasone augments myogenic differentiation in C2C12 cells¹⁵, potentially by enhancing and modifying IGF-1 and/or insulin-receptor tyrosine-kinase-mediated signalling¹⁶. I/D treatment, even with insulin doses of 1–10 nM, induced a rapid and dramatic fusion of myocytes into extensive rope-like networks and, like IGF-1, induced myotube hypertrophy, a switch to glycolytic metabolism and activation of calcineurin, responses that could all be prevented by cyclosporin A (Fig. 3a–c). Unlike IGF-1, insulin increases only near-membrane Ca^{2+} concentrations¹⁷. Consistent with this finding, I/D treatment did not significantly change $[\text{Ca}^{2+}]_i$ ($\Delta[\text{Ca}^{2+}]_i = +8 \pm 16 \text{ nM}$). Nevertheless, I/D treatment mobilized Ca^{2+} , as the increase in CnPP activity was prevented by pretreatment with BAPTA-AM (Fig. 2b).

We next investigated the timing of IGF-1-induced myotube hypertrophy and its inhibition by cyclosporin A. As reported⁴, the hypertrophy observed in IGF-1-transfected C2C12 cells could be mimicked by addition of recombinant IGF-1 (250 ng ml^{-1}) to untransfected cells. This effect was associated with Ca^{2+} mobilization ($\Delta[\text{Ca}^{2+}]_i = +72 \pm 14 \text{ nM}$, $P < 0.001$ versus untreated cells), and with activation of calcineurin, which could be prevented by BAPTA-AM (Fig. 2b). However, hypertrophy of C2C12 cells was observed only if IGF-1 was added before or at the time of differentiation, but not one day after. Similarly, cyclosporin A inhibited myotube hypertrophy of IGF-1-transfected cells only when added before or at the time of differentiation, but not after. These results identify a critical window for induction of hypertrophy, early in the transition from myoblast (satellite cell) to post-mitotic myocyte.

In cardiac hypertrophy, the apparent target of calcineurin-mediated dephosphorylation is the transcription factor NF-AT3 (ref. 7). At least three NF-AT isoforms are expressed in mouse and human skeletal muscle cells, and with mobilization of intracellular Ca^{2+} they translocate to the nucleus^{18,19}. Therefore we investigated NF-AT expression and cellular localization by immunoblotting of cytosolic and nuclear extracts prepared from control and IGF-1-transfected C2C12 myoblasts and myotubes at 1, 3 and 5 days post-differentiation. At least three 90K–116K NF-AT species were identified with an antibody specific for NF-ATc1, but not with antibodies for NF-ATc2 (NF-AT1) or NF-ATc4 (NF-AT3). However, all NF-ATc1 species identified were present in both the cytosolic and nuclear extracts from myoblasts and myotubes at all stages, and from both control and IGF-1-transfected cells (data not shown). Given that there may be a critical window for induction of hypertrophy and subsequent restoration of NF-AT distribution, we tested whether acute activation of calcineurin could result in NF-AT dephosphorylation and nuclear translocation. In the presence of serum-free medium, both the Ca^{2+} ionophore A23187 (1.5 μM) and recombinant IGF-1 (250 ng ml^{-1}) induced nuclear translocation of NF-ATc1 isoforms (Fig. 3d), as well as the accumulation of smaller species of lower relative molecular mass, as observed for the dephosphorylated forms of NF-AT identified in other systems²⁰. However, the acute IGF-1 response was transient, peaking at 30 min but no longer evident at 60 min (Fig. 3d).

Finally, we investigated the *in vivo* effects of IGF-1 by directly injecting a plasmid encoding IGF-1 into latissimus dorsi muscle (LDM) of rats; the contralateral LDM, which received vector only, served as control. About 10% of skeletal muscle cells in the immediate vicinity (2 mm radius) of the injection sites were transformed, as seen by co-expressed green fluorescent protein, and expression of IGF-1 messenger RNA persisted for at least three months (Fig. 4a, b). When collected two months after IGF-1 injection, LDM showed not only increased calcineurin activity and lactate levels (Fig. 4b, c), but also an increased number of myofibres with central nuclei (Fig. 4d), a hallmark of new muscle formation due to satellite-cell activation². This is consistent with the

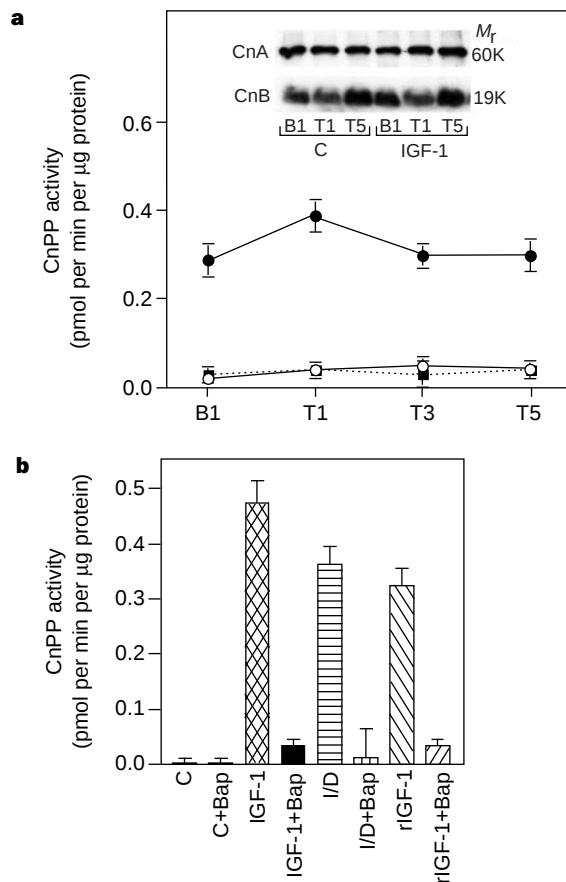


Figure 2 Calcineurin phosphatase (CnPP) activity. **a**, CnPP activity in control (circles), IGF-1 (filled circles) or IGF-1-plus-CsA(squares)-treated myoblasts (B1, one day after plating) and myotubes (T1–T5; one, three or five days after changing to differentiation medium). Inset, CnA and CnB expression (immunoblot analysis; equal protein loading per lane) in control and IGF-1 treated myoblasts and myotubes. **b**, CnPP activity in control (C), IGF-1-transfected (IGF-1), insulin (1 μM)-plus-dexamethasone(I/D)-treated and recombinant IGF-1-treated (rIGF-1) cells, in the absence or presence of BAPTA-AM (Bap), at 6 h post-differentiation. BAPTA-AM (200 μM ; Calbiochem) was added for 1 h in serum-free medium, and the cells were then collected for analysis.

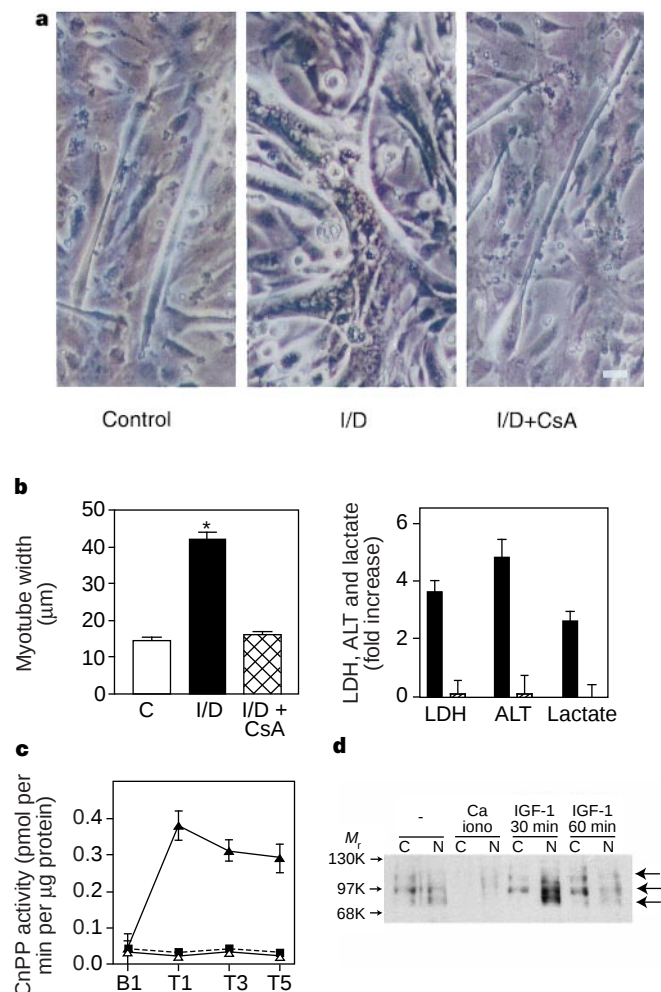


Figure 3 Induction and calcineurin-dependence of C2C12 cell myotube hypertrophy resulting from I/D treatment. **a**, Morphology of control (C), insulin- (1 μM)-plus-dexamethasone (2.5 μM) (I/D)-, and I/D-plus-CsA (1 μM)-treated myotubes, three days post-differentiation (scale bar, 20 μm). **b**, Myotube width, and ALT, LDH and lactate levels, in C-, I/D- and I/D-plus-CsA-treated myotubes, five days post-differentiation (**P* < 0.001). **c**, Calcineurin phosphatase (CnPP) activity in C(triangles)-, I/D(filled triangles)- or I/D-plus-CsA(squares)-treated myoblasts or myotubes. **d**, Immunoblots of cytosolic (C) and nuclear (N) extracts (equal protein loading) from native C2C12 myotubes using an anti-NF-ATc1 (NF-ATc1) antibody. Cells at 6 h post-differentiation treated with calcium ionophore (iono, 1.5 μM) for 4 h, or recombinant IGF-1 (250 ng ml⁻¹) for the times indicated. Arrows indicate the migration of NF-ATc1 species.

contention that maintenance of type IIb fibres with IGF-1/AAV (adeno-associated virus)-treatment of muscle, which prevents age-related loss of muscle mass and strength², and generation of hypertrophy in muscles subjected to chronic work overload²¹, involves fusion of activated satellite cells with, rather than direct effects on, existing myofibres. Although calcineurin activation has also been implicated in promoting a slow rather than fast-twitch phenotype in untreated muscle at steady state¹⁸, our findings and those of ref. 2 support a contrasting role in an IGF-1-dependent adaptive/regenerative process that enhances glycolytic myofibres. These different outcomes may relate to activation of additional pathways by IGF-1 in parallel to calcineurin signalling.

Based on our findings, we propose a model whereby skeletal muscle hypertrophy and a switch in myofibre phenotype is initiated by a stimulus that causes Ca²⁺ mobilization. This activates calcineurin and results in NF-ATc1 dephosphorylation and nuclear translocation. The resulting transcriptionally competent NF-ATc1 species then initiate a programme of gene expression that leads to

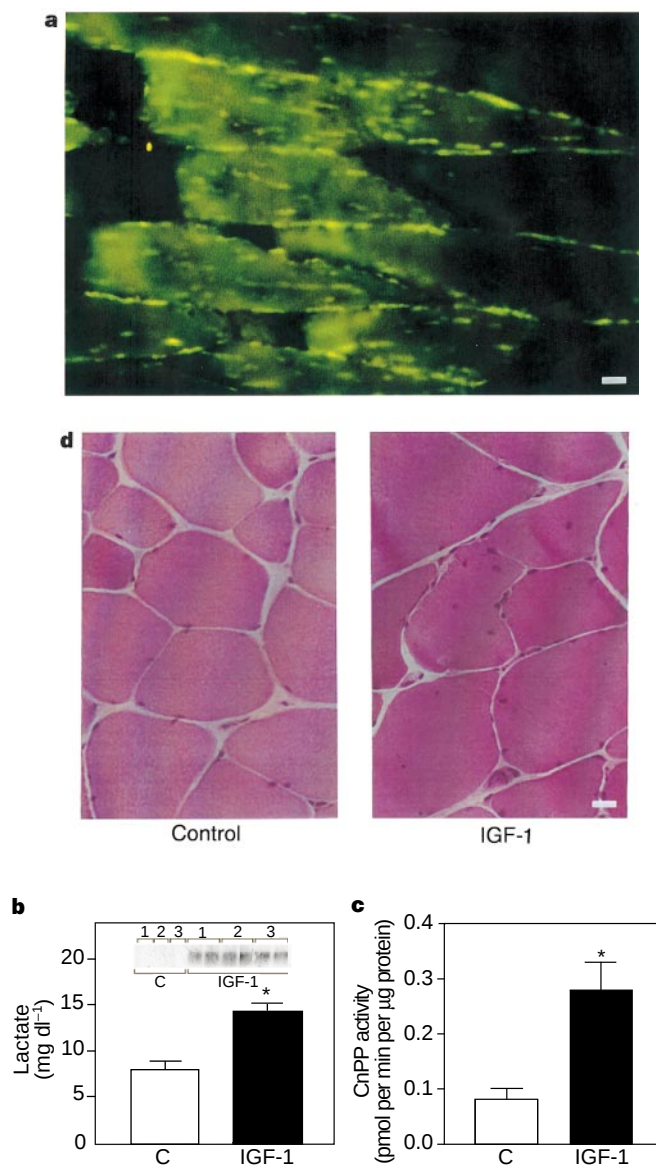


Figure 4 IGF-1-induced transformation of rat latissimus dorsi muscle (LDM). **a**, Green fluorescence of IGF-1-transformed LDM (scale bar, 10 μm). **b**, Lactate levels of control (C) and IGF-1-transformed (IGF-1) LDM, **P* < 0.001. Inset, northern blot showing IGF-1 mRNA in control and pIGF-1/IRES/GFP20-treated LDM at 1, 2 and 3 months post-injection (IGF-1 mRNA from one control and two IGF-1-injected LDM samples are shown at each time point; equal RNA loading). **c**, Calcineurin phosphatase (CnPP) activity in control and IGF-1-transformed LDM (**P* < 0.001). **d**, Haematoxylin-and-eosin stained control and IGF-1-injected LDM. Central nuclei were apparent in 3–10% of IGF-1 myofibres, but were not detected in control myofibres (0%, *P* < 0.001) (scale bar, 10 μm).

hypertrophy. Once activated, this pathway becomes resistant to subsequent calcineurin inhibition. If similar regulatory mechanisms operate in cardiomyocytes, this may partly explain the failure of cyclosporin A or FK506 to cause regression of hypertrophy, despite inhibition of calcineurin⁹. As in cardiomyocytes, transactivation of hypertrophy genes in skeletal muscle cells probably involves interaction with a GATA transcription factor because GATA2, which is weakly expressed in untreated L6E9 skeletal muscle cells, increases in response to IGF-1 and is complexed with dephosphorylated NF-ATc1 and calcineurin in the nucleus²².

In immune cells, calcineurin is sensitive to both transient calcium spikes and tonic increases in [Ca²⁺]_i (ref. 23). However, persistent nuclear translocation of NF-AT occurs only with tonic increases²³. Furthermore, in baby hamster kidney (BHK) cells, sustained

nuclear translocation of NF-AT4 requires binding of activated calcineurin, which masks nuclear export sequences and thereby prevents the rapid removal of NF-AT4 from the nucleus²⁴. Binding of activated calcineurin may also be required to induce a conformational change in NF-AT that is necessary for transcriptional activity²⁵, either as a direct consequence of binding or as a result of further dephosphorylation of NF-AT. Our finding that skeletal muscle cells also use calcineurin and NF-AT in the induction of hypertrophy indicates that this signalling pathway provides a mechanism for a variety of cells to interpret differences in second messenger amplitude and duration, induced by various external stimuli. How this discrimination occurs is not known, but it may relate to the kinetics of pathway deactivation that directly involves nuclear export of NF-AT, as in immune cells^{23,24,26}, to homeostatic regulation of nuclear transport, as indicated by our studies in muscle cells, or to complementary pathways using distinct Ca^{2+} -discriminating mechanisms²⁷. □

Methods

Plasmid construction. A plasmid (pIGF-1/IRES/GFP20), allowing expression of IGF-1 under the control of an elongation factor-1 α promoter, was prepared and used to stably transfect C2C12 cells by co-transfection with pcDNA3.neo, as described⁴. pIGF-1/IRES/GFP20 contains an internal ribosome entry site and complementary DNA for the green fluorescent protein variant GFP20 downstream of the IGF-1 cDNA, allowing co-expression of GFP20. A control vector was identical but lacked the IGF-1 cDNA.

Cell lines, transfection and culture. We maintained stably transfected clones in Dulbecco's modified Eagle's medium (DMEM) containing 20% fetal calf serum and Geneticin (200 $\mu\text{g ml}^{-1}$). To induce differentiation, the medium was changed to DMEM plus 2% horse serum and Geneticin (200 $\mu\text{g ml}^{-1}$). GFP20 fluorescence was assessed by epifluorescence using a fluorescein isothiocyanate filter (Zeiss OxioLab). Cells were grown in the absence or presence of cyclosporin A, LY294002 (Sigma), rapamycin (ICN Laboratories) or PD98059 (New England Biolabs), as indicated.

Evaluation of protein and DNA synthesis and anaerobic glycolysis. We determined the fractional protein synthesis rate ($[\text{L-}[2,6\text{-}^3\text{H}]\text{-phenylalanine}$ (Amersham) uptake), DNA synthesis ($[\text{L-}[^3\text{H}]\text{thymidine}$ (Amersham) incorporation), and anaerobic glycolysis (lactate dehydrogenase and alanine aminotransferase activities, and lactate levels) as described⁴.

Histological analysis. Myotube tube morphology was determined after fixation in 10% methanol and staining with 5% Giemsa. Myotube area and width were determined by quantitative morphometry using Leica Q500MC image analysis software as described⁴. For each treatment group, 100 myotubes were analysed from five randomly selected high-power ($\times 40$ magnification) fields.

Calcineurin phosphatase (CnPP) activity. C2C12 cells were washed in PBS and lysed by freeze/thawing in 50 μl of buffer containing 50 mM Tris-HCl, pH 7.5, 0.1 mM EGTA, 1 mM EDTA, 0.5 mM dithiothreitol, 0.5 mM phenylmethylsulphonylfluoride, and 5 μg each of leupeptin, aprotinin and soybean trypsin inhibitor. Extracts were cleared by centrifugation at 10,000g for 10 min at 4°C. CnPP activity in resulting supernatant fractions was determined in the presence of 0.1 mM Ca^{2+} and 100 nM calmodulin as described²⁸ and calculated from the difference in ^{32}P released in the absence and presence of a calcineurin-specific inhibitory peptide (peptide 412)²⁸.

Immunoblot analysis. For calcineurin A (CnA) and B (CnB) subunits, we prepared whole-cell extracts from C2C12 cells by adding 200 μl lysis buffer (2% SDS, 10 mM Tris-HCl, pH 7.4) per 10-cm dish after three initial washes with PBS. For NF-ATs, cells were fractionated into a cytosolic extract and nuclear pellet as described²³, apart from adding 1 mM DTT to the hypotonic buffer. Nuclear proteins were then released from the nuclear pellet to give a nuclear extract as described²⁹. After boiling for 5 min, the supernatants or cytosolic and nuclear extracts were fractionated by SDS-PAGE (8% for detection of CnA and NF-ATs; 12% for CnB). Immunoblotting followed standard procedures with anti-CnA (1:3,000 final dilution), anti-CnB (1:2,000), anti-NF-ATc1 (NF-AT2) (1:500 final dilution), anti-NF-ATc2 (NF-AT1) (1:500 final dilution) or anti-NF-ATc4 (NF-AT3) (1:500 final dilution) (Santa Cruz Biotechnology) primary antibodies, followed by a sheep anti-mouse HRP-tagged (1:1,000; Amersham)

or a donkey anti-goat HRP-tagged secondary antibody (1:2,000; Amersham), and resolution by ECL chemiluminescence (New England Life Science).

Intracellular calcium measurements in C2C12 cells. These were made as described³⁰. Briefly, cells were loaded with 5 μM indo-1-AM for 15 min at room temperature, and $[\text{Ca}^{2+}]_i$ was calculated from the 400/505 nm fluorescence ratio corrected for background and autofluorescence. Indo-1 ratio signals were converted to $[\text{Ca}^{2+}]_i$ using a calibration curve.

In vivo IGF-1 transformation. Rats (Wistar male, $n = 8$) were anaesthetized with 5% halothane and the latissimus dorsi muscles (LDMs) exposed by a midline incision. pIGF-1/IRES/GFP20 (100 μg) was injected into one LDM and the control vector (100 μg) into the contralateral LDM by using a 25-gauge needle and 20 injections per cm^2 of muscle. The incision was then closed with surgical skin clips. After recovery, animals were maintained on standard rat chow and water. At the times indicated, animals were killed and LDMs removed for RNA extraction and northern blot analysis⁴, calcineurin phosphatase activity and histology. All studies were performed in accordance with NH and MRC guidelines and with the approval of the St Vincent's Hospital Animal Ethics Committee.

Data analysis. Comparisons between treatments were made using paired or unpaired t -tests or χ^2 analysis, with $P < 0.05$ being considered statistically significant. All data are expressed as the mean $\pm$ s.e.m. (error bars).

Received 26 April; accepted 1 June 1999.

- Pette, D. & Staron, R. S. Mammalian skeletal muscle fiber type transitions. *Int. Rev. Cytology* **170**, 143–223 (1997).
- Barton-Davis, E. R. *et al.* Viral mediated expression of insulin-like growth factor 1 blocks the aging-related loss of skeletal muscle function. *Proc. Natl Acad. Sci. USA* **95**, 15603–15607 (1998).
- Florini, J. R., Ewton, D. Z. & Coolican, S. A. Growth hormone and the insulin-like growth factor system in myogenesis. *Endocrine Rev.* **17**, 481–517 (1996).
- Semsarian, C., Suttrave, P., Richmond, D. R. & Graham, R. M. IGF-induced myotube hypertrophy associated with an increase in anaerobic glycolysis in a clonal skeletal muscle cell model. *Biochem. J.* **339**, 443–451 (1999).
- Tesch, P. A., Thorsson, A. & Kaiser, P. Muscle capillary supply and fiber type characteristics in weight and power lifters. *J. Appl. Physiol. Respirat. Environ. Exercise Physiol.* **56**, 35–38 (1984).
- Suttrave, P., Kelly, A. M. & Hughes, S. H. *ski* can cause selective growth of skeletal muscle in transgenic mice. *Genes Dev.* **4**, 1462–1472 (1990).
- Molkentin, J. D. *et al.* A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* **93**, 215–228 (1998).
- Luo, Z., Shyu, K.-G., Gualberto, A. & Walsh, K. Calcineurin inhibitors and cardiac hypertrophy. *Nature Med.* **4**, 1092–1093 (1998).
- Zhang, W. *et al.* Failure of calcineurin inhibitors to prevent pressure-overload left ventricular hypertrophy in rats. *Circ. Res.* **84**, 722–728 (1999).
- Renganathan, M., Messi, M. L. & Delbono, O. Overexpression of IGF-1 exclusively in skeletal muscle prevents age-related decline in the number of dihydropyridine receptors. *J. Biol. Chem.* **273**, 28845–28851 (1998).
- Wang, Z. M., Messi, M. L., Renganathan, M. & Delbono, O. Insulin-like growth factor-1 enhances rat skeletal muscle charge movement and L-type Ca^{2+} channel gene expression. *J. Physiol.* **516**, 331–341 (1999).
- Kazaki, M., Nie, L., Shibata, H. & Kojima, I. Activation of a calcium-permeable cation channel CD20 expressed in Balb/C 3T3 cells by insulin-like growth factor-1. *J. Biol. Chem.* **272**, 4964–4969 (1997).
- Coolican, S. A. *et al.* The mitogenic and myogenic actions of insulin-like growth factors utilize distinct signalling pathways. *J. Biol. Chem.* **272**, 6653–6662 (1997).
- Musaro, A. & Rosenthal, N. Maturation of the myogenic program is induced by post-mitotic expression of IGF-1. *Mol. Cell. Biol.* **19**, 3115–3124 (1999).
- Montano, M. M. & Lim, R. W. Glucocorticoid effects on the skeletal muscle differentiation program: analysis of clonal proliferation, morphological differentiation and the expression of muscle-specific regulatory genes. *Endocr. Res.* **23**, 37–57 (1997).
- Giorgino, F. & Smith, R. J. Dexamethasone enhances insulin-like growth factor-1 effects on skeletal muscle cell proliferation. *J. Clin. Invest.* **96**, 1473–1483 (1995).
- Bruton, J. D., Katz, A. & Westerblad, H. Insulin increase near-membrane but not global Ca^{2+} in isolated skeletal muscle. *Proc. Natl Acad. Sci. USA* **96**, 3281–3286 (1999).
- Chin, E. R. *et al.* A calcineurin-dependent transcriptional pathway controls skeletal muscle fiber type. *Genes Dev.* **12**, 2499–2509 (1998).
- Abbott, K. L. *et al.* Activation and cellular localization of the cyclosporine A-sensitive transcription factor NF-AT in skeletal muscle cells. *Mol. Biol. Cell* **9**, 2905–2916 (1998).
- Beals, C. R. *et al.* Nuclear export of NF-ATc enhanced by glycogen synthase kinase-3. *Science* **275**, 1930–1933 (1997).
- Rosenblatt, J. D. & Parry, D. J. Gamma irradiation prevents compensatory hypertrophy of overloaded mouse extensor digitorum longus muscle. *J. Appl. Physiol.* **73**, 2538–2543 (1992).
- Musaro, A., McCullagh, K. J. A., Naya, F. J., Olson, E. N. & Rosenthal, N. IGF-1 induces skeletal myocyte hypertrophy through calcineurin in association with GATA-2 and NF-ATc1. *Nature* **400**, 581–585 (1999).
- Dolmetsch, R. E., Lewis, R. S., Goodnow, C. C. & Healy, J. I. Differential activation of transcription factors induced by Ca^{2+} response amplitude and duration. *Nature* **386**, 855–858 (1997).
- Zhu, J. & McKeon, G. NF-AT activation requires suppression of Crm1-dependent export by calcineurin. *Nature* **398**, 256–260 (1999).
- Hogan, P. G. & Rao, A. Modification by nuclear export? *Nature* **398**, 200–201 (1999).
- Crabtree, G. R. Generic signals and specific outcomes: signaling through Ca^{2+} calcineurin, and NF-AT. *Cell* **96**, 611–614 (1999).
- De Koninck, P. & Schulman, H. Sensitivity of CaM kinase II to the frequency of Ca^{2+} oscillations. *Science* **279**, 227–230 (1998).
- Fruman, D. A. *et al.* Measurement of calcineurin phosphatase activity in cell extracts. *Methods* **9**, 146–154 (1996).
- Lyakh, L., Ghosh, P. & Rice, N. R. Expression of NFAT-family proteins in normal human T cells. *Mol. Cell. Biol.* **17**, 2465–2484 (1997).

30. Ju, Y.-K. & Allen, D. G. Intracellular calcium and Na-Ca²⁺ exchange current in isolated toad pacemaker cells. *J. Physiol.* **508**, 153–166 (1998).

Acknowledgements. Supported in part by funds donated in memory of Diana, Princess of Wales. We thank P. Poronnik and C. Lee for IGF-1 expression vector reagents; N. Rosenthal for sharing unpublished data; P. Gunning for C2C12 cells; M. Fujioka (Fujisawa) for supplying FK506; C. Klee for the 'Free Calcium' program; K. Walsh, E. Hardeman, A. Gualberto and P. Suttrave for helpful discussions; and E. Martin and N. Groves for preparation for this manuscript. This work was supported by NH and MRC Medical Research Scholarships (to C.S. and T.Y.).

Correspondence and requests for materials should be addressed to R.M.G. (e-mail: b.graham@victorhang.unsw.edu.au).

IGF-1 induces skeletal myocyte hypertrophy through calcineurin in association with GATA-2 and NF-ATc1

Antonio Musarò*, Karl J. A. McCullagh*, Francisco J. Naya†, Eric N. Olson† & Nadia Rosenthal*

* Cardiovascular Research Center, Massachusetts General Hospital—East, 149 13th Street, Charlestown, Massachusetts 02129, USA

† Department of Molecular Biology and Oncology, the University of Texas Southwestern Medical Center, 6000 Harry Hines Blvd, Dallas, Texas 75225, USA

Localized synthesis of insulin-like growth factors (IGFs) has been broadly implicated in skeletal muscle growth, hypertrophy and regeneration¹. Virally delivered IGF-1 genes induce local skeletal muscle hypertrophy and attenuate age-related skeletal muscle atrophy, restoring and improving muscle mass and strength in mice². Here we show that the molecular pathways underlying the hypertrophic action of IGF-1 in skeletal muscle are similar to those responsible for cardiac hypertrophy. Transfected IGF-1 gene expression in postmitotic skeletal myocytes activates calcineurin-mediated calcium signalling by inducing calcineurin transcripts and nuclear localization of calcineurin protein. Expression of activated calcineurin mimics the effects of IGF-1, whereas expression of a dominant-negative calcineurin mutant or addition of cyclosporin, a calcineurin inhibitor, represses myocyte differentiation and hypertrophy. Either IGF-1 or activated calcineurin induces expression of the transcription factor GATA-2, which accumulates in a subset of myocyte nuclei, where it associates with calcineurin and a specific dephosphorylated isoform of the transcription factor NF-ATc1. Thus, IGF-1 induces calcineurin-mediated signalling and activation of GATA-2, a marker of skeletal muscle hypertrophy, which cooperates with selected NF-ATc isoforms to activate gene expression programs.

We have previously generated a skeletal-muscle-cell model, L6MLC/IGF-1, in which hypertrophy is induced by post mitotic expression of the non-circulating IGF-1 muscle isoform³ in rat L6E9 myocytes⁴. It has been proposed^{5,6} that the calcineurin signal-transduction pathway, which induces cardiac hypertrophy⁷, might also control skeletal muscle differentiation. Calcineurin is a ubiquitous calcium-activated serine phosphatase, a heterotrimeric complex comprising a catalytic subunit (CnA), a tightly associated regulatory subunit (CnB) and calmodulin, a calcium sensor⁸.

To determine whether calcineurin signalling participates in IGF-1-mediated hypertrophy, we analysed the expression patterns of the CnA and CnB subunits in parental L6E9 and in L6MLC/IGF-1 cultures during differentiation (Fig. 1A). Transcripts encoding the CnA subunit were not detected in L6E9 cells under proliferating (growth medium) or differentiating (differentiation medium) culture conditions. L6MLC/IGF-1 myocytes accumulated abundant CnA transcripts, one of which (4.5 kb) has previously been described in skeletal muscle⁹, whereas a novel transcript (1.7 kb) was induced by IGF-1 expression. CnB transcripts were expressed at

similar levels in both L6E9 and L6MLC/IGF-1 cultures (Fig. 1A), indicating that the gene for CnB is not activated by IGF-1. In contrast, we detected significant amounts of CnA protein in L6E9 myocytes, which were further increased in L6MLC/IGF-1 myocytes (Fig. 1B), as well as in C2C12 skeletal myocytes (data not shown). Whereas calcineurin localized to both cytoplasm and nuclei of differentiated L6E9 myocytes, it became predominantly nuclear in L6MLC/IGF-1 hypertrophic myocytes, which are larger and have unusual nuclear clusters⁴.

A functional role for calcineurin in the hypertrophic response was tested in L6E9 cultures stably transfected with a calcium-independent form of activated CnA¹⁰ driven by a myogenin promoter (L6Myog/CnA). Differentiated L6Myog/CnA myocytes displayed a pronounced hypertrophic phenotype (Fig. 2A) typical of differentiated L6MLC/IGF-1 cultures⁴. Moreover, transient transfection of L6MLC/IGF-1 myoblasts with a dominant-negative CnA mutant tagged with haemagglutinin (CnA-DN-HA)¹¹ blocked subsequent differentiation, characterized by the mutually exclusive appearance of HA-tagged CnA-DN and muscle myosin (Fig. 2B).

Cyclosporin A effectively blocks calcineurin phosphatase activity, both in cardiac tissue⁷ and in primary skeletal muscle cultures⁶. Proliferating L6E9 or L6MLC/IGF-1 myoblasts treated with cyclosporin could not differentiate, as analysed by MyHC expression (data not shown). To determine whether later stages of IGF-1-mediated hypertrophy were also dependent on the calcineurin pathway, cyclosporin was added to quiescent L6E9 and L6MLC/IGF-1 cultures,

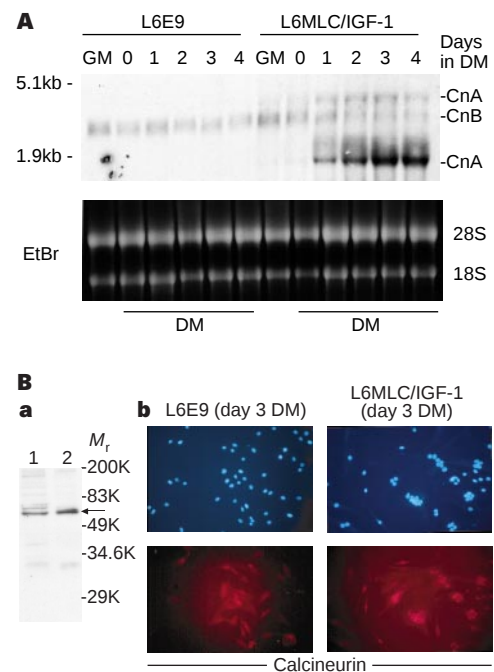


Figure 1 Calcineurin expression and subcellular localization are responsive to postmitotic IGF-1 expression in myogenic cell cultures. **A**, Effects of MLC/IGF-1 expression on CnA and CnB transcripts in differentiating L6E9 cultures. L6E9 or L6MLC/IGF-1 myoblasts⁴ were maintained in GM or shifted to DM for the number of days indicated. Northern-blot analysis revealed a characteristic 4.0-kb species previously detected in skeletal muscle⁹, and a new ~1.7-kb species that increased in response to IGF-1-induced hypertrophy. EtBr, ethidium bromide staining; 28S and 18S indicate ribosomal RNA markers. Effects of MLC/IGF-1 expression on calcineurin protein levels and subcellular distribution in L6E9 myocytes. **a**, Western-blot analysis of 61K CnA protein in differentiated L6E9 (lane 1) or L6MLC/IGF-1 (lane 2) myocytes harvested at day 3 (DM), using an anti-CnA antibody. *M_r*, relative molecular mass. **b**, Immunofluorescent analysis of calcineurin in the same cultures. Hoechst staining (top, blue) and anti-CnA antibody (bottom, red) reveals the predominantly nuclear localization of calcineurin in the L6MLC/IGF-1 hypertrophic myocytes.

before or after the myogenic differentiation programme was activated, and they were analysed for MyHC expression after 5 days. Both L6E9 and L6MLC/IGF-1 cultures treated with cyclosporin at day 0 failed to differentiate (Fig. 2C), as demonstrated by the absence of MyHC protein. The addition of cyclosporin at day 2, when MLC/IGF-1 expression was already activated, did not block MyHC accumulation (Fig. 2C), indicating that IGF-1 might counteract the effects of cyclosporin through other pathways. Similar results have been obtained by the administration of cyclosporin to the myogenic C2C12 cell line¹². The narrow window of susceptibility to cyclosporin-mediated inhibition of differentiation argues against general toxicity of the drug in these experiments. In L6MLC/IGF-1 cultures, hypertrophy (but not differentiation) was compromised by cyclosporin at day 2 (Fig. 2C), revealing a specific

effect of cyclosporin on the hypertrophic phenotype, and indicating that continuous calcineurin activity may be necessary for the maintenance of IGF-1-induced hypertrophy.

How does IGF-1 activate calcineurin? In cardiac muscle, a sustained elevation in intracellular $[Ca^{2+}]$ elicits a hypertrophic response through the calmodulin–calcineurin pathway^{7,13}. In rat skeletal myocytes, IGF-1 enhances charge movement by activating L-type calcium-channel gene expression¹⁴, in a manner similar to that of insulin, which increases near-membrane $[Ca^{2+}]$ in isolated skeletal muscle fibres¹⁵. Increasing intracellular $[Ca^{2+}]$ in L6MLC/IGF-1 cultures by treatment with calcium ionophore for 2 h immediately after serum withdrawal significantly increased hypertrophy in the L6MLC/IGF-1 cultures 4 days later, even when the cells were seeded at very low density (Fig. 2D). In contrast, similar

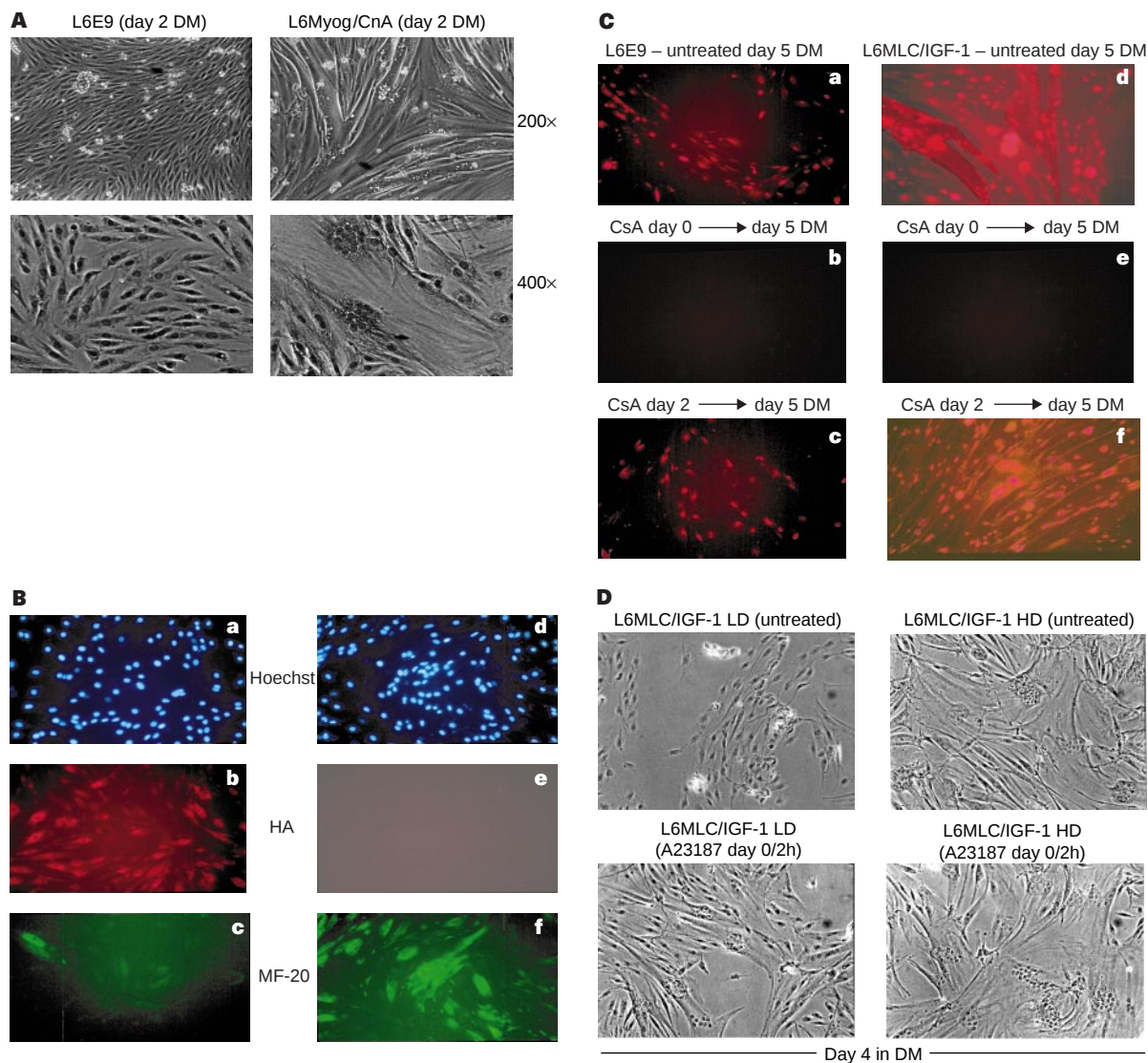


Figure 2 Perturbation of calcineurin action affects hypertrophy in myogenic cell cultures. **A**, Hypertrophic response to activated CnA expression in post-mitotic L6E9 myocytes. Parental L6E9 and L6Myog/CnA cultures were differentiated for two days in DM. Phase-contrast photographs of representative fields from the two cultures are shown at 200x and 400x original magnification. **B**, Suppression of myogenic differentiation by constitutive expression of CnA-DN-HA in transiently transfected L6MLC/IGF-1 cultures¹¹. Transfectants were switched to DM for 2 days and the CnA-DN-HA mutant (red) and muscle myosin (green) were analysed by immunofluorescence with anti-HA and MF20 antibodies, respectively. **a–c**, HA-positive transfectants; **d–f**, Untransfected cells from the same plate. Myocyte hypertrophy, characterized by nuclear rings (**d**) and muscle myosin expression (**f**) was blocked in those cells transfected with the CnA-DN-

HA plasmid (compare **b** and **c**). **C**, Cyclosporin (CsA) blocks myogenic differentiation and hypertrophy. L6E9 (**a–c**) and L6MLC/IGF-1 (**d–f**) cultures were differentiated in parallel with 5 μ M CsA, added either at day 0 (**b**, **e**) or day 2 (**c**, **f**) after the cells were switched to DM. **a**, **d**, Untreated cultures. CsA added at day 0 blocked myogenic differentiation in both lines by day 5, as visualized by MF20 antibody against muscle myosin (**b**, **e**). CsA added at day 2 did not block myosin expression but reduced hypertrophy in the L6MLC/IGF-1 myocytes (compare **d** and **f**). Similar experiments using 3 μ M CsA gave identical results (not shown). **D**, Calcium ionophore increases myocytes hypertrophy. L6MLC/IGF-1 myoblasts were treated with 1.5 μ M calcium ionophore A23187 at 50% (LD) or 90% (HD) of confluence (day 0) for 2 h, and then switched to DM and analysed at day 4.

treatment of untransfected C2C12 and L6E9 cultures with calcium ionophore resulted in marked cell death, with reduced numbers of myocytes after 3 days (data not shown). Although other studies have shown that IGF-1 is directly responsible for elevating intracellular calcium^{12,14,15}, our results indicate that high intracellular $[Ca^{2+}]$ is lethal to differentiating myocytes, unless other pathways induced by IGF-1 are subsequently activated.

The hypertrophic effects of calcineurin activity in cardiomyocytes have been associated with the GATA-4 transcription factor, which is also critical for proper heart development¹⁶. GATA factors have not previously been implicated in skeletal muscle gene regulation. In C2C12 skeletal muscle cultures, we have detected GATA-2 transcripts in myoblasts that declined during differentiation, whereas low levels of GATA-2 protein accumulated in differentiated myocytes (data not shown). In contrast, GATA-2 transcripts were undetectable in L6E9 cultures, but were induced in L6MLC/IGF-1 myocytes (Fig. 3A). GATA-2 activation by IGF-1 was mediated at least partly by calcineurin, because the addition of cyclosporin to quiescent L6MLC/IGF-1 myoblasts reduced subsequent GATA-2 activation on differentiation (Fig. 3B: compare lanes 2 and 3). GATA-2 transcripts were also induced in L6Myog/CnA myocytes expressing activated calcineurin (Fig. 3B: compare lanes 1 and 4). Consistent with its transcript distribution, GATA-2 protein was not detected by immunofluorescence analysis in L6E9 or L6MLC/IGF-1 proliferative myoblasts (data not shown), nor in differentiated L6E9 cultures (Fig. 3C, a). Abundant GATA-2 protein was localized to the nuclei in hypertrophic L6MLC/IGF-1 myocytes, but it was absent

from the nuclei of non-hypertrophic myocytes in less densely seeded regions of the same culture (Fig. 3C, b). GATA-2 protein was also induced in the nuclear clusters of L6Myog/CnA hypertrophic myocytes (Fig. 3D, b), confirming its activation by the calcineurin pathway.

In cardiac hypertrophy, GATA-4 forms nuclear complexes with another transcription factor, NF-ATc4 (NF-AT3), a target of calcineurin phosphatase activity in a variety of cells. Accumulation of GATA-2 exclusively in hypertrophic nuclei of L6MLC/IGF-1 and L6Myog/CnA cultures led us to investigate the potential role of associated NF-ATc factors in the hypertrophic response. Skeletal myocytes express NF-ATc1 (NF-AT2/c), NF-ATc2 (NF-AT1/p) and NF-ATc3 (NF-ATx/4)⁶, some of which undergo nuclear translocation in the presence of activated calcineurin^{5,6}. GATA-2 and NF-ATc proteins co-localized to the same subset of nuclei within a hypertrophic L6MLC/IGF-1 myocyte (Fig. 4A). In co-immunoprecipitations, calmodulin-associated calcineurin (Fig. 4B; CnA panel, lanes 1 and 2) complexed specifically with NF-ATc2 in both L6E9 and L6MLC/IGF-1 myocytes (Fig. 4B; NF-ATc2 panel, lanes 1 and 2). NF-ATc2 protein in these complexes was phosphorylated, although less so in the hypertrophic cultures (Fig. 4B; phosphoserine panel, lanes 1 and 2). These data support a model in which partial dephosphorylation of NF-ATc by calcineurin uncovers the nuclear-localization signal but does not confer full transcriptional activity¹⁷. In contrast, GATA-2 was associated exclusively with complexes containing calcineurin and a single isoform of NF-ATc1 (Fig. 4B; lane 3, CnA and NF-ATc1 panels) that was extensively

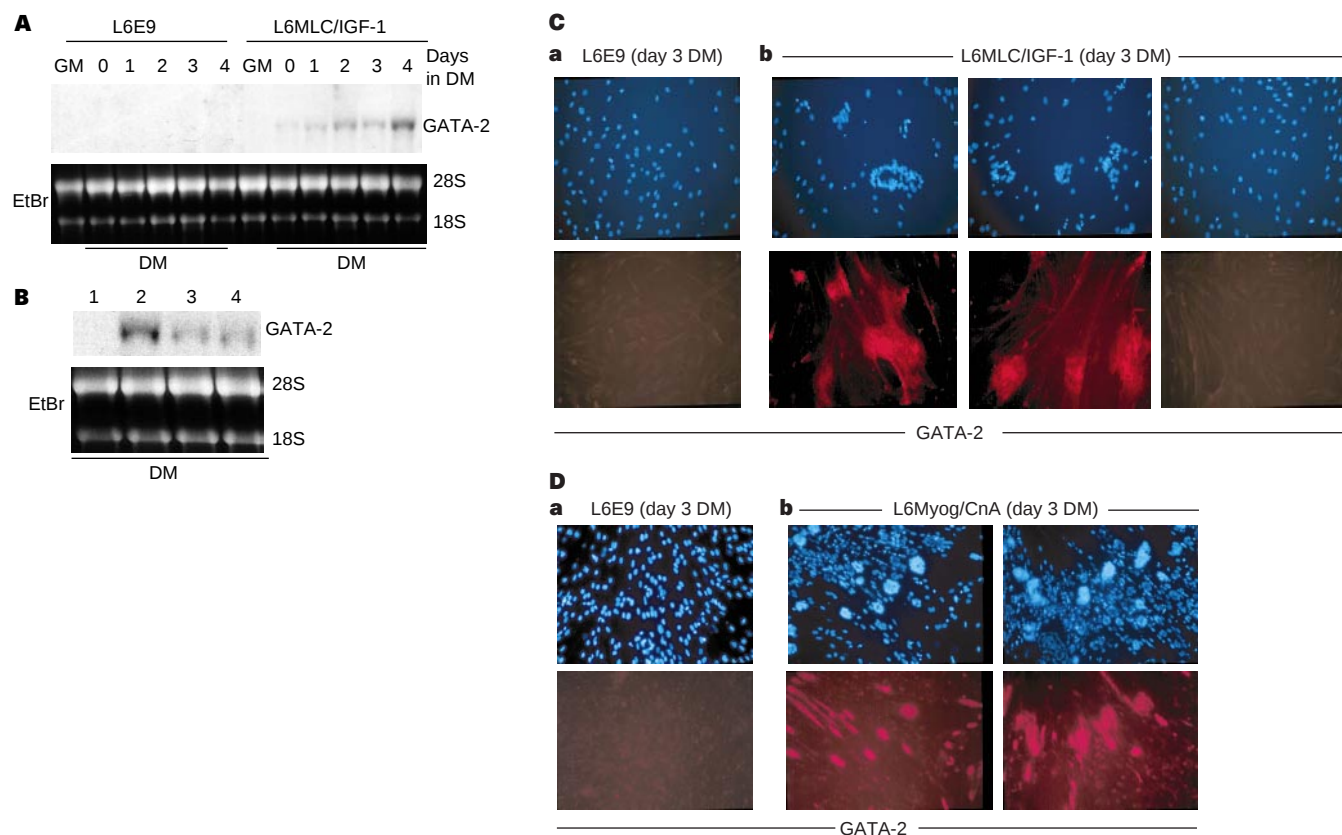


Figure 3 Activation of GATA-2 expression in hypertrophic myocytes. **A**, Northern-blot analysis of GATA-2 transcripts in L6E9 or L6MLC/IGF-1 cultures maintained in GM or differentiated in DM for up to 4 days. **B**, Northern-blot analysis of GATA-2 transcripts in L6E9 (lane 1), L6MLC/IGF-1 (lane 2), L6MLC/IGF-1 + 3 μ M CsA added at day 0 (lane 3), and L6Myog/CnA cultures after 2 days in DM (lane 4). **C**, Nuclear localization of GATA-2 protein exclusively in L6MLC/IGF-1 hypertrophic myocytes. L6E9 (**a**) and L6MLC/IGF-1 (**b**) cultures were differentiated in DM for 3 days, then analysed for GATA-2 protein (red). Top panels, nuclear Hoechst stain (blue). Of the three fields shown for L6MLC/IGF-1 cultures, the two left-hand

fields show GATA-2 localized to hypertrophic myocytes, characterized by nuclear clusters; the right-hand field shows a less densely seeded area from the same plate devoid of hypertrophic myocytes. **D**, Subcellular localization of GATA-2 protein in nuclei of L6Myog/CnA hypertrophic myocytes. L6E9 (**a**) and L6Myog/CnA (**b**) cultures were differentiated in DM for 3 days, then analysed for GATA-2 protein by immunofluorescence (red). Top, nuclear Hoechst stain (blue). Note GATA-2 localization to hypertrophic myocytes, characterized by nuclear rings, exclusively in the L6Myog/CnA cultures.

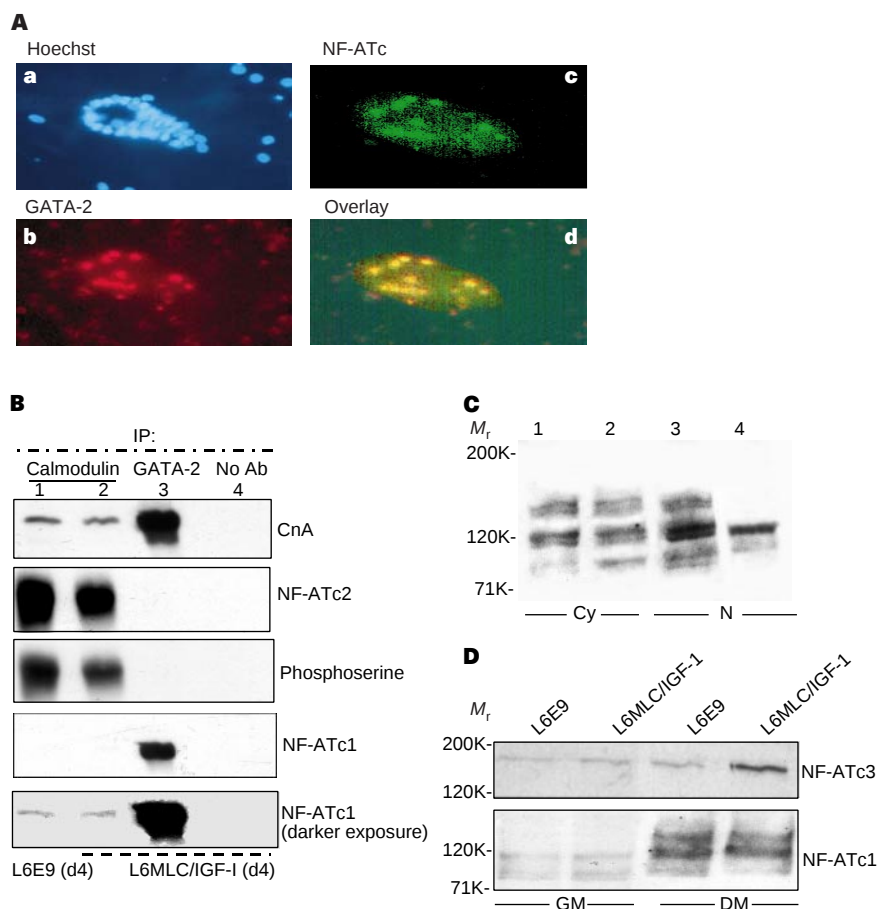


Figure 4 GATA-2 interacts with NF-ATc1 to induce myocyte hypertrophy. **A**, GATA-2 and NF-ATc colocalize to the same nuclear subset within a hypertrophic myocyte. **a–d**, Differentiated L6MLC/IGF-1 myocytes were immunostained with **b**, anti-GATA-2 mouse monoclonal, and **c**, anti-NF-AT polyclonal antibodies, or **a**, were nuclear (Hoechst) stained. **d**, Overlay panel shows the merged image between NF-ATc and GATA-2 immunofluorescence. **B**, NF-ATc1 co-precipitates preferentially with GATA-2. Differentiated L6E9 and L6MLC/IGF-1 lysates were incubated with either calmodulin (lanes 1 and 2) or GATA-2 (lane 3) antibodies. Western-blot analysis of co-immunoprecipitates (IP) was done with anti-calceineurin (CnA) and anti-NF-ATc2 antibodies, then the blot was stripped and

reprobed with anti-NF-ATc1 and anti-phosphoserine antibodies. Lane 4 shows L6MLC/IGF-1 lysates without antibodies. **C**, Subcellular distribution of NF-ATc1 isoforms. Western-blot analysis of nuclear (N) and cytoplasmic (Cy) extracts from differentiated L6E9 (lanes 1, 3) and L6MLC/IGF-1 (lanes 2, 4) cultures. Doublets of NF-ATc1 isoforms reflect differences in phosphorylation. Predominant NF-ATc1 isoform in lane 4 is identical to that complexed with GATA-2 in **B**. **D**, NF-ATc1 and NF-ATc3 are differentially induced in L6E9 and L6MLC/IGF-1 cultures. Western blots of total cellular extracts from myoblasts (GM) or myocytes (DM), analysed with specific NF-ATc antibodies as indicated.

dephosphorylated (Fig. 4B; lane 3, phosphoserine panel) and was predominantly expressed in hypertrophic myocytes (Fig. 4C). This indicates that maintenance of calcineurin-mediated dephosphorylation may not be dependent on continuous activation by calmodulin, although low levels of the same dephosphorylated NF-ATc1 isoform were detected in calmodulin-associated complexes (Fig. 4B; NF-ATc1 panel, lanes 1 and 2). Western blot analysis of fractionated myocyte extracts revealed three distinct NF-ATc1 isoforms in L6E9 nuclear and cytoplasmic fractions (Fig. 4C: compare lanes 1 and 3), whereas one dephosphorylated isoform preferentially accumulated in L6MLC/IGF-1 nuclear fractions (lane 4). NF-ATc3 was not detected in any of the co-immunoprecipitates in Fig. 4B (not shown), although it was present in total cellular extracts from L6E9 and L6MLC/IGF-1 myoblasts or differentiated myocytes, induced to higher levels in L6MLC/IGF-1 hypertrophic myocytes (Fig. 4D). In contrast, small amounts of multiple NF-ATc1 isoforms in L6E9 and L6MLC/IGF-1 myoblasts increased in both L6E9 and L6MLC/IGF-1 myocytes (Fig. 4D). These results underscore the isoform-dependent nature of NF-ATc regulation in skeletal muscle cultures, and highlight the specificity of the NF-ATc1–GATA-2 interaction. Although the specific role of NF-ATc1 in skeletal muscles has not been determined, it has been implicated in

morphogenesis of the cardiac valves and septum during heart development^{18,19}.

The physiological relevance of our findings is supported by the phenotype of MLC/IGF-1 transgenic mice, which develop pronounced hypertrophic skeletal muscles with persistent activation of both CnA and GATA-2 expression, implicating a similar signalling cascade operating *in vivo* (manuscript in preparation). However, additional pathways may contribute to muscle hypertrophy. Signal transduction mediated by phosphatidylinositol is required for initiation of the hypertrophic response to IGF-1 (ref. 4). IGF-1 also affects glucose transport²⁰, which is upregulated in stretch-induced hypertrophy in L6 myocytes²¹. The action of MEF2 transcription factors is required for Glut-4 gene transcription²². Although MEF2c expression is activated in L6MLC/IGF-1 cultures⁴, it is not induced in the L6Myog/CnA line (data not shown). Additional transcription factors in IGF-1-stimulated muscle cultures^{4,23,24} may collaborate with GATA-2 and NF-ATc1 to induce hypertrophy *in vivo*. Other cofactors and substrates may modulate calcineurin signalling in response to tonic motor-nerve activity, which sustains high intracellular calcium levels characteristic of slow twitch muscle fibres⁵. Further resolution of these pathways has important implications for IGF-1 as a gene-therapeutic agent in

ageing and diseased muscle². Our results, together with those from previous studies⁷, show that skeletal and cardiac muscle cells share elements of a common signalling system in which calcineurin regulates muscle hypertrophy by controlling combinatorial associations of NF-AT and GATA transcription factors. □

Methods

Cell culture and transfection. The myogenic parental cell line L6E9^{25,26} and the stable clone L6MLC/IGF-1⁴ were maintained in growth medium (GM) consisting of Dulbecco's modified Eagle's medium (DMEM) supplemented with 20% fetal bovine serum. For stable transfection, L6E9 culture cells were transfected with a calcium-independent form of activated CnA under the control of the myogenin promoter (Myo 1565) and with puromycin-selectable-vector (pPUR, Clontech) (10:1), as described⁴. We isolated twelve drug-resistant clones (L6Myog/CnA) using cloning rings, and characterized them morphologically and by northern blot analysis for CnA expression. The morphologies of high- and low-expressing L6Myog/CnA clones were equivalent.

For transient transfections, L6MLC/IGF-1 cells were transfected with a calcineurin-dominant negative-HA-tagged plasmid (CnA-DN-HA)¹¹. To activate myogenic differentiation, myoblasts were switched to differentiation medium (DM)(DMEM + 1% BSA) and analysed by double immunofluorescence for CnA-DN-HA and MyHC expression after 2 days in DM.

Calcium-ionophore treatment. Cells were grown as described above and treated with 1.5 μ M calcium ionophore A23187 (or control vehicle DMSO) at 50% (low density) or 90% (high density) of confluence (day 0) for 2 h, then washed with PBS and switched to DM. After 4 days in DM, cells were analysed morphologically.

Treatment with cyclosporin A. Cells were grown as described above and treated with either 3 or 5 μ M of cyclosporin A (CsA) as outlined in the test, at different times in the differentiation program: 80% confluence (day 0), and day 2 (d2) in DM. Cells treated at day 0 or day 2 were continuously exposed to CsA until they were collected at day 5. Cells treated during GM with CsA in ethanol were switched to DM without CnA inhibitor (or ethanol as control vehicle) when they were at 80% confluence. At day 5, cells were analysed by northern blotting or immunofluorescence analysis.

RNA preparation and northern blot analysis. Total RNA from L6E9, L6MLC/IGF-1 and L6Myog/CnA cultures was obtained by RNA-TRIZOL extraction (Gibco-BRL); 15 μ g of total RNA was fractionated by electrophoresis on 1.3% agarose gels and hybridized as described⁴.

Immunofluorescence analysis. Differentiated L6E9, L6MLC/IGF-1 and L6Myog/CnA cultures were fixed in 4% paraformaldehyde and processed as described⁴. The nuclei of myogenic cells were visualized using Hoechst staining. We used the following antibodies: a monoclonal MF-20 antibody (Hybridoma Bank); a monoclonal antibody against calcineurin (Transduction Labs); a monoclonal antibody against GATA-2 (Santa Cruz); a goat polyclonal against NF-ATc (Santa Cruz); and a rabbit polyclonal against HA (Santa Cruz).

Immunoprecipitation. L6E9 and L6MLC/IGF-1 myotubes were extracted with 1% Triton X-100 buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 10 mM sodium pyrophosphate, 10 mM NaF and 1% Triton X-100, supplemented with 20 mM sodium pyrophosphate, 10 mM NaF, 1 mM sodium orthovanadate, 10 μ g ml⁻¹ aprotinin, 20 μ g ml⁻¹ soybean trypsin inhibitor and leupeptin, and 2 mM phenylmethylsulphonyl fluoride) and supplemented with 1 mM CaCl₂ (ref. 27). Samples containing 700 μ g of total protein extract were immunoprecipitated with rabbit polyclonal anti-calmodulin antibody (Zymed), or with mouse monoclonal anti-GATA-2 antibody (Santa Cruz) overnight at 4°C, then incubated with A/G-Sepharose beads at 4°C for 1 h. Beads were washed three times in wash buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 5 mM sodium pyrophosphate, 10% glycerol, 1% Triton X-100, and protease inhibitors supplemented with 1 mM CaCl₂) and once in 10 mM HEPES, pH 7.4, and then boiled in Laemmli reducing buffer. The samples were analysed by 6% SDS-PAGE, followed by western blotting with anti-calcineurin, anti-NF-ATc1 (provided by A. Rao), anti-NF-ATc1 (provided by G. Crabtree)²⁸ and anti-phosphoserine (Zymed) antibodies.

Extracts and subcellular fractionation. Differentiated L6E9 and L6MLC/IGF-1 cultures were lysed in hypotonic buffer (5 mM NaCl, 20 mM HEPES, pH 7.5)²⁹ containing protease (20 μ g ml⁻¹ soybean trypsin inhibitor and leupeptin,

and 2 mM phenylmethylsulphonyl fluoride) and phosphatase (10 mM NaF, 1 mM sodium orthovanadate) inhibitors and 0.4% Nonidet-P40. Nuclei were separated by centrifugation at 800g and cytoplasmic extracts (supernatant) were added to boiling SDS-PAGE reducing sample buffer. Nuclei were rinsed in hypotonic buffer with inhibitors and lysed in boiling sample buffer. Extracts were resolved by SDS-PAGE, transferred to nitrocellulose and probed with NF-ATc1 antibody.

Western blot analysis. Differentiated L6E9 and L6MLC/IGF-1 cultures were lysed in lysis buffer (25 mM HEPES pH 7.2, 100 mM NaCl, 2 mM MgCl₂, CaCl₂, 0.5% Triton X-100, 1 mM PMSE, and 1 μ g ml⁻¹ each of leupeptin, antipain, aprotinin and pepstatin)³⁰. Proteins were resolved by SDS-PAGE, transferred to nitrocellulose and probed with calcineurin and NF-ATc antibodies.

Received 20 April; accepted 14 June 1999.

- Florini, J., Ewton, D. & Magri, K. Hormones, growth factors, and myogenic differentiation. *Annu. Rev. Physiol.* **53**, 201–216 (1991).
- Barton-Davis, E., Shoturma, D. I., Musaro, A., Rosenthal, N. & Sweeney, H. L. Viral mediated expression of IGF-1 blocks the aging-related loss of skeletal muscle function. *Proc. Natl Acad. Sci. USA* **95**, 15603 (1998).
- Shimatsu, A. & Rotwein, P. Mosaic evolution of the insulin-like growth factors: organization, sequence, and expression of the rat insulin-like growth factor I gene. *J. Biol. Chem.* **262**, 7894–7900 (1987).
- Musaro, A. & Rosenthal, N. Maturation of the myogenic program is induced by post-mitotic expression of IGF-I. *Mol. Cell. Biol.* **19**, 3115–3124 (1999).
- Chin, E. *et al.* A calcineurin-dependent transcriptional pathway controls skeletal muscle fiber type. *Genes Dev.* **12**, 2499–2509 (1998).
- Abbot, K., Friday, B. B., Thaloer, D., Murphy, T. J. & Pavlath, G. K. Activation and cellular localization of the cyclosporine A-sensitive transcription factor NF-AT in skeletal muscle cells. *Mol. Biol. Cell* **9**, 2905–2916 (1998).
- Molkentin, J. *et al.* A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* **93**, 215–228 (1998).
- Crabtree, G. Generic signals and specific outcomes: signaling through Ca²⁺, calcineurin and NF-AT. *Cell* **96**, 611–614 (1999).
- Kincaid, R. L. *et al.* Cloning and characterization of molecular isoforms of the catalytic subunit of calcineurin using nonisotopic methods. *J. Biol. Chem.* **265**, 11312–11319 (1990).
- Manalan, A. & Klee, C. Activation of calcineurin by limited proteolysis. *Proc. Natl Acad. Sci. USA* **80**, 4291–4295 (1983).
- Muramatsu, T. & Kincaid, R. L. Inhibition of NF-AT signal transduction events by dominant-negative form of calcineurin. *Biochem. Biophys. Res. Commun.* **218**, 466–472 (1996).
- Semsarian, C. *et al.* Skeletal muscle hypertrophy is mediated by a Ca²⁺-dependent calcineurin-dependent signalling pathway. *Nature* **400**, 576–581 (1999).
- Gruver, C. L., DeMayo, F., Goldstein, M. A. & Means, A. R. Targeted developmental overexpression of calmodulin induces proliferative and hypertrophic growth of cardiomyocytes in transgenic mice. *Endocrinology* **133**, 376–388 (1993).
- Wang, Z., Messi, M. L., Renganathan, M. & Delbono, O. Insulin-like growth factor-I enhances rat skeletal muscle charge movement and L-type Ca²⁺ channel gene expression. *J. Physiol.* **516**, 331–341 (1999).
- Bruton, J., Katz, A. & Westerblad, H. Insulin increases near-membrane but not global Ca²⁺ in isolated skeletal muscle. *Proc. Natl Acad. Sci. USA* **96**, 3281–3286 (1999).
- Molkentin, J. D., Lin, Q., Duncan, S. A. & Olson, E. N. Requirement of the transcriptional factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev.* **11**, 1061–1072 (1997).
- Hogan, P. G. & Rao, A. Modification by nuclear export? *Nature* **398**, 200–201 (1999).
- de la Pompa, J. L. *et al.* Role of the NF-ATc transcription factor in morphogenesis of cardiac valves and septum. *Nature* **392**, 182–186 (1998).
- Ranger, A. M. *et al.* The transcription factor NF-ATc is essential for cardiac valve formation. *Nature* **392**, 186–190 (1998).
- Wilson, C. M., Mitsumoto, Y., Maher, F. & Klip, A. Regulation of cell surface GLUT1, GLUT3, and GLUT4 insulin and IGF-1 in L6 myotubes. *FEBS Lett.* **10**, 19–22 (1995).
- Mitsumoto, Y., Downey, G. P. & Klip, A. Stimulation of glucose transport in L6 muscle cells by long-term intermittent stretch-relaxation. *FEBS Lett.* **301**, 94–98 (1992).
- Thai, M. V., Gurusamy, S., Cao, K. T., Pessin, J. E. & Olson, A. L. Myocyte enhancer factor 2 (MEF2)-binding site is required for GLUT4 gene expression in transgenic mice. Regulation of MEF2 DNA binding activity in insulin-deficient diabetes. *J. Biol. Chem.* **273**, 14285–14292 (1998).
- Florini, J. R., Ewton, D. Z. & Magri, K. A. Insulin-like growth factor-1 stimulates terminal myogenic differentiation by induction of myogenin gene expression. *Mol. Endocrinol.* **5**, 718–724 (1991).
- Engert, J., Berglund, E. B. & Rosenthal, N. Proliferation precedes differentiation in IGF-1 stimulated myogenesis. *J. Cell Biol.* **135**, 431–440 (1996).
- Nadal-Ginard, B. Commitment, fusion and biochemical differentiation of a myogenic cell line in the absence of DNA synthesis. *Cell* **15**, 855–864 (1996).
- Yaffe, D. Retention of differentiation potentialities during prolonged cultivation of myogenic cells. *Proc. Natl Acad. Sci. USA* **61**, 477–483 (1968).
- Loh, C. *et al.* Calcineurin binds the transcription factor NFAT1 and reversibly regulates its activity. *J. Biol. Chem.* **271**, 10884–10891 (1996).
- Northrop, J. P. *et al.* NF-AT components define a family of transcription factors targeted in T-cell activation. *Nature* **369**, 497–502 (1994).
- Dolmetsch, R. E., Lewis, R. S., Goodnow, C. C. & Healy, J. I. Differential activation of transcription factors induced by Ca²⁺ response amplitude and duration. *Nature* **368**, 855–858 (1997).
- Clifton, N. A., Fiorentino, D. F. & Crabtree, G. R. Molecular analysis of the interaction of calcineurin with drug-immunophilin complexes. *J. Biol. Chem.* **269**, 26431–26437 (1994).

Acknowledgements. We thank L. Sweeney, F. McKeon, C. Marks, and members of the Rosenthal laboratory for helpful discussion; A. Rao and G. Crabtree for advice and gifts of NF-AT antibodies; M. McGrew, C. Semsarian and R. Graham for sharing unpublished data; and J. Florini for his original concept and continued insights into the properties of IGF. The study was supported by grants to N.R. from the NIA and NASA, to A.M. from the MDA, and to E.N.O. from the NIH. E.N. was supported by a NIH postdoctoral fellowship.

Correspondence and requests for materials should be addressed to N.R.

Inner-arm dynein c of *Chlamydomonas* flagella is a single-headed processive motor

Hitoshi Sakakibara*, Hiroaki Kojima*, Yukako Sakai*, Eisaku Katayama† & Kazuhiro Oiwa*

* Kansai Advanced Research Center, Communications Research Laboratory, Kobe 651-2401, Japan

† Department of Fine Morphology, Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo 108-8639, Japan

Axonemal dyneins are force-generating ATPases that produce movement of eukaryotic cilia and flagella^{1,2}. Several studies indicate that inner-arm dyneins mainly produce bending moments in flagella^{3,4} and that these motors have inherent oscillations in force and motility^{5–8}. Processive motors such as kinesins have high duty ratios of attached to total ATPase cycle (attached plus detached) times⁹ compared to sliding motors such as myosin¹⁰. Here we provide evidence that subspecies-c, a single-headed axonemal inner-arm dynein, is processive but has a low duty ratio. Ultrastructurally it is similar to other dyneins^{11–14}, with a single globular head, long stem and a slender stalk that attaches to microtubules. *In vitro* studies of microtubules sliding over surfaces coated with subspecies-c at low densities (measured by single-molecule fluorescence) show that a single molecule is sufficient to move a microtubule more than $1\ \mu\text{m}$ at $0.7\ \mu\text{m s}^{-1}$. When many motors interact the velocity is $5.1\ \mu\text{m s}^{-1}$, fitting a duty ratio of 0.14. Using optical trap nanometry, we show that beads carrying a single subspecies-c motor move processively along the microtubules in 8-nm steps but slip backwards under high loads. These results indicate that dynein subspecies-c functions in a very different way from conventional motor proteins, and has properties that could produce self-oscillation *in vivo*.

Axonemal dyneins exist in many isoforms: in *Chlamydomonas* flagellar axonemes, the outer-arm dyneins have three heavy chains and the inner-arm dyneins have eight^{2,3,15}. These dynein heavy chains have distinct functions, each of which is crucial for proper axonemal function³. The eight distinct inner-arm dynein heavy chains are organized with various intermediate and light chains into seven different molecular complexes: one two-headed and six single-headed isoforms^{11,15}. We separated the inner-arm dyneins into seven subspecies, termed a to g, by high-performance liquid chromatography (Fig. 1a), and have used only purified subspecies-c. This has a single globular head, 12 nm in diameter, and a stalk, analogous to the structural organization of other phylogenetically and functionally diverse dyneins^{11–14,16} (Fig. 1b, c). The head domain of subspecies-c consists of several globular subdomains encircling a hollow central core, similar to cytoplasmic dynein¹⁴ (Fig. 1b, d). Subspecies-c incubated with microtubules without ATP appeared as globular structures with stems aligned obliquely to the microtubule but without direct contact of either the globular structure or the stem with the microtubule (Fig. 1e). These observations indicate that the microtubule-binding site of axonemal subspecies-c may be segregated away from the globular head and located at the tip of a stalk, as indicated for cytoplasmic dynein¹⁶.

For functional studies of subspecies-c, we used the *in vitro* motility assay in which microtubules glide on glass surfaces coated with dynein¹⁵. We determined the effects on motility of the density of subspecies-c adsorbed on glass surfaces, by incubation with solutions of progressively lower concentrations of subspecies-c. As subspecies-c was distributed evenly in solution and did not

aggregate (Fig. 1d), this procedure quantitatively reduced the number of subspecies-c molecules able to interact with individual microtubules. We determined the surface density of active subspecies-c molecules from each starting concentration directly, using a single-molecule technique developed for myosin^{17,18}: the fluorescent ATP analogue, Cy3-EDA-ATP, was infused and the fluorescence spots per unit area were counted. This technique showed that the number of subspecies-c molecules bound to the surface decreased in proportion to their solution concentration and that they were evenly distributed. Dynein surface densities measured by this technique were 10-fold less than estimated by assuming that all dynein molecules in the solution were bound to the surface. We used the surface densities determined by this single-molecule technique, as this directly measures only active subspecies-c molecules. To assay motor function at low surface densities¹⁹, we measured the rate at which microtubules came out of bulk solution, bound and moved more than $0.8\ \mu\text{m}$ over surfaces coated with subspecies-c; a displacement of less than $0.8\ \mu\text{m}$ is difficult to distinguish from Brownian motion. Addition of 0.05% methylcellulose and

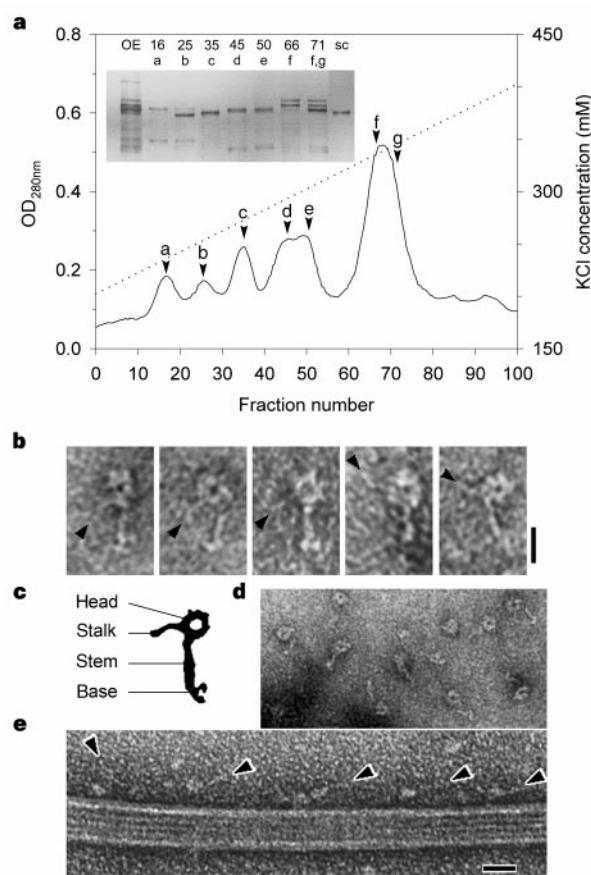


Figure 1 Biochemical and ultrastructural characterization of subspecies-c. **a**, HPLC fractionation of high-salt *oda1* axoneme extracts (dotted line is KCl concentration). Inset, enlargement of the high-molecular-mass region of SDS-PAGE of separation peaks of the dynein subspecies (arrowheads); subspecies-c has a relative molecular mass of 600K¹⁵. Figures indicate fraction number and subspecies (a–g). Lanes OE and sc are, respectively, *oda1* extract and subspecies-c after further purification. **b**, Subspecies-c conformers visualized by negative-staining electron microscopy in the absence of nucleotide. Arrowheads indicate the globular tip of the stalk. Scale bar, 10 nm. **c**, Drawing of subspecies-c morphology with components labelled^{11,12}. **d**, Survey view of purified subspecies-c, showing an even distribution of homogeneous particles without aggregation. **e**, Association of subspecies-c with a single microtubule in the absence of ATP. Heads of subspecies-c molecules are aligned along the microtubule but at a distance. Arrowheads indicate the stems of each molecule, pointing away from the microtubule. Scale bar, 20 nm.

1 mg ml⁻¹ bovine serum albumin to the solution allowed us to observe attachment at very low densities of subspecies-c: attachment and movement of microtubules were observed even at densities of one molecule per 100 μm^2 . The log-log plot of rate of microtubule attachment against subspecies-c surface density is well fitted by the relationship for a single subspecies-c molecule being sufficient for attachment and motility of microtubules^{19,20} (Fig. 2a).

Moving microtubules observed at subspecies-c densities below one molecule per 10 μm^2 rotated erratically about a vertical axis through a fixed point on the surface, at which a single motor molecule is presumably located¹⁹, while still progressing (Fig. 2b, c). The maximal angle of microtubule rotation around such nodal points was $\sim 17^\circ$. When its trailing end reached this nodal point, the microtubule dissociated from the surface and diffused back into solution. At high subspecies-c densities, however, microtubules exhibited negligible rotation when moving, presumably because they were constrained laterally by attachment of two or more subspecies-c molecules. Both these patterns of movement

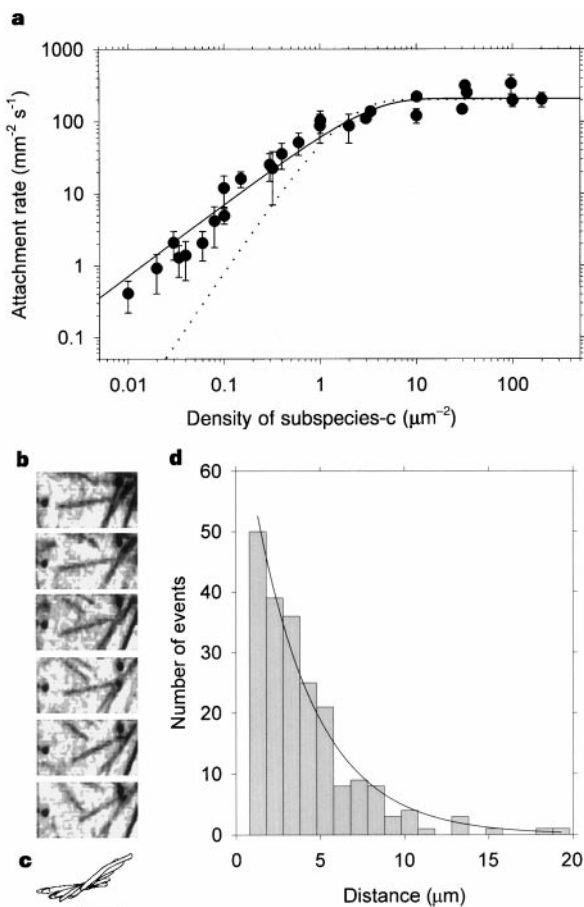


Figure 2 Microtubule motion *in vitro* driven by subspecies-c. **a**, Dependence of microtubule attachment rate, number observed per mm² per s, on the density of subspecies-c on glass surfaces with 0.05% methylcellulose. Solid curve is the fit to the Poisson model for a single molecule of subspecies-c sufficing to move a microtubule: $k(\rho) = C_0(1 - e^{-A\rho})$, where $k(\rho)$ is attachment rate at subspecies-c density ρ and A is the product of microtubule length and twice the reach of subspecies-c; these values are $C_0 = 208 \text{ mm}^{-2} \text{ s}^{-1}$, $A = 0.34 \mu\text{m}^2$. Dotted line is fit to the Poisson model for two molecules, $k(\rho) = C_0(1 - e^{-A\rho} - A\rho e^{-A\rho})$. Error bars are s.d. **b**, **c**, Video sequence and tracings of images showing thermally driven rotation of a microtubule around a nodal point while also moving progressively (subspecies-c density: 1 molecule per 100 μm^2). Tracings were obtained from successive video images at 1-s intervals. **d**, Distribution of distances travelled on a nodal point by microtubules before detachment. Data were pooled for densities of 1 to 10 molecules per 100 μm^2 . Solid curve: exponential fit with length constant 3.6 μm . ATP concentration, 0.5 mM.

were similar to those observed for microtubules moving over kinesin-coated glass surfaces¹⁹. The direction of longitudinal movement while pivoting was determined with polarity-marked microtubules²¹; the marked end (minus) was almost always trailing (20 out of 23 microtubules). However, microtubules with poorly-defined polarity in the sample ($\sim 10\%$) could appear to move in the opposite direction. These observations indicate that subspecies-c moves toward the minus end of microtubules and therefore that microtubules were not driven by contaminating axonemal kinesin²².

We measured the distances travelled by microtubules over surfaces coated with low densities of subspecies-c (less than one molecule per 10 μm^2) after binding to a nodal point. Microtubules greater than 10 μm in length that moved more than 0.8 μm were followed over subspecies-c densities of 1–10 molecules per 100 μm^2 . Most microtubules (88%) that landed on a nodal point moved a distance shorter than their length, were released from the nodal point and diffused away from the surface back into solution. The remaining microtubules that landed on a nodal point moved right to their end before being released. No microtubules were observed moving more than their length within this density range. A histogram of the distances travelled by microtubules shows a single exponential distribution with a distance constant of 3.6 μm (Fig. 2d) about three times as long as that for kinesin⁹.

The relationship between the velocity of microtubule movement and surface density of subspecies-c was determined with 0.5 mM ATP, close to saturating concentration (see Fig. 3). On surfaces coated with high densities of subspecies-c (>30 molecules per μm^2), microtubules (mean length 10 μm) moved continuously over distances greater than their length at velocities of 5.1 $\mu\text{m s}^{-1}$. As the surface density of subspecies-c was reduced, the sliding velocity decreased markedly. Below surface densities of 1 molecule per 10 μm^2 , most microtubules ($>95\%$) pivoted around a single nodal point and moved progressively at velocities of 0.7 $\mu\text{m s}^{-1}$, independent of their length, which ranged from 3 to 25 μm . According to the model of ref. 10, the average velocity of microtubules at high subspecies-c densities is limited by the step-size d divided by the power-stroke duration (t_s), during which a subspecies-c molecule propels the microtubule the distance d ; the velocity of microtubule movement at low subspecies-c densities is limited by d divided by the total ATPase cycle time (t_c), as, under these conditions, the displacement of d is generated at an average frequency of $1/t_c$. Thus, the ratio of the average velocity at very low subspecies-c densities ($V_{\min}$) to that at high densities ($V_{\max}$) gives an estimate of the duty ratio, assuming that the 'step-size' of subspecies-c at high densities is the same as that at low densities. Fits to

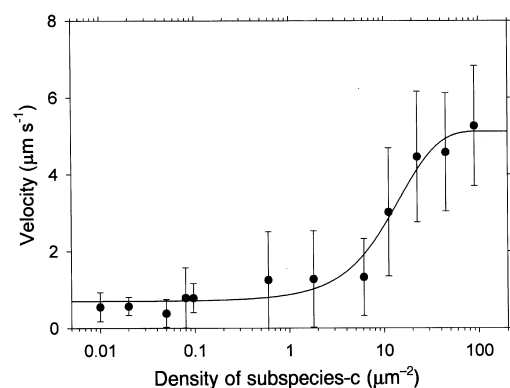


Figure 3 Relationship between velocity of microtubule translation and subspecies-c surface density. Error bars indicate s.d. Solid curve is a fit to $V_{\max}(1 - e^{-A\rho})/(1 - e^{-A\rho_f})$ in which $V_{\max}$, ρ , f and A are, respectively, the maximal velocity, the density of subspecies-c, the duty ratio and the product of microtubule length and twice the reach of subspecies-c; values: $V_{\max} = 5.1 \mu\text{m s}^{-1}$, $V_{\min} = 0.7 \mu\text{m s}^{-1}$, $f = V_{\min}/V_{\max} = 0.14$, $A = 0.53 \mu\text{m}^2$.

the velocity data gave a duty ratio of 0.14, much smaller than that (>0.99) of the well-defined processive motor kinesin⁹.

To elucidate the mechanism of processive motion at small duty ratios using more precise measurements, we made a bead assay using an optical trap²³. Polystyrene beads coated with various densities of subspecies-c molecules were prepared by mixing subspecies-c solutions with beads at various molar ratios. A diffusing bead was captured with the optical trap and brought into contact with a single microtubule bound to the glass surface. The bead would often attach to the microtubule, showing markedly reduced Brownian motion. After a few seconds, the trap laser was blocked by a shutter, whereupon beads that had failed to bind to a microtubule diffused away. These beads could be recaptured and re-tried. A reduction in Brownian motion by 50% was the criterion for the establishment of binding of a bead to a microtubule.

The relative frequency of binding depends on the number of subspecies-c molecules carried by a bead. Assuming that one active subspecies-c molecule, attached to a bead with a geometry that is favourable for interaction, suffices to bind the bead to a microtubule, the fraction of beads bound to microtubules should follow Poisson statistics²⁴. Our data are fitted well by the single-molecule probability $P(n) = (1 - e^{-n/\lambda})$, where n is the number ratio of subspecies-c molecules mixed with beads (see Fig. 4a, with $\lambda = 480 \pm 73$). λ is the binding efficiency: there is a 50% probability that a bead will have one active subspecies-c molecule bound when beads and subspecies-c molecules are mixed at a number ratio of $0.69\lambda(331)$. Measurements of force and movement were made using beads having less than 50% probability of binding and hence movement, so the Poisson probability that a bead carried two or more active subspecies-c molecules was less than 0.15. Assuming a

random distribution of subspecies-c molecules over the bead surface and allowing a 50-nm reach for the subspecies-c molecule's tether¹¹ (see Fig. 1b), the probability that two or more subspecies-c molecules were able to interact simultaneously with a microtubule is less than 0.01.

The trap stiffness was chosen to be as large as possible to reduce Brownian motion of beads while not exerting forces so large that a subspecies-c molecule was restricted from producing its full displacement. For these measurements a compliant optical trap, 0.017 pN nm^{-1} , was used. When a trapped bead was brought into contact with a microtubule, it attached and executed transient movements along the direction of the microtubule (Fig. 4b), such that several runs occurred as a cluster. In such clusters, beads progressed for various periods and then were drawn rapidly back to the trap centre, whereupon they rebound and started moving again. Multiple cycles of attachment, movement and release were observed with the same bead. The average distance moved was $69 \pm 25 \text{ nm}$, corresponding to a trap force of $1.2 \pm 0.4 \text{ pN}$ (mean $\pm$ s.d., $n = 77$). These progressive displacements were mostly towards the microtubules' minus ends, as confirmed by polarity-marking (30 out of 32 experiments). Displacements of beads perpendicular to the microtubule axis were noisy owing to the low trap stiffness: the root-mean-square displacement of beads was 16 nm .

In records of bead displacement, thermal noise decreased as the beads developed greater force, indicating that the stiffness of a subspecies-c molecule increases as it strains away from its microtubule-binding site (Fig. 4b, c). This phenomenon, a departure from Hooke's law, was also reported for kinesin-coated beads²⁴. At forces from 1 to 2 pN, the step movement of beads was clearly

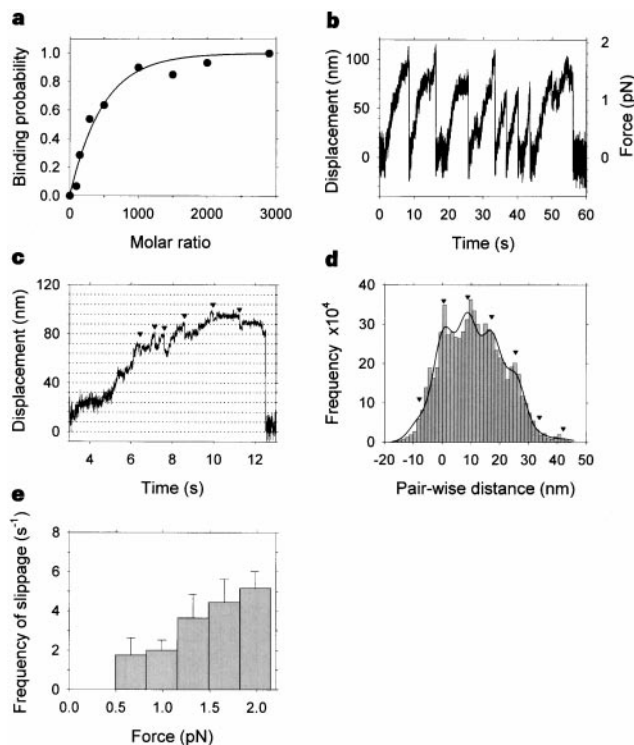


Figure 4 Optical trap nanometry of subspecies-c-coated beads. **a**, Fraction of beads binding and moving as a function of the molar mixing ratio n of subspecies-c molecules to beads at incubation. The fitted curve is $(1 - e^{-n/\lambda})$, with the efficiency of binding $\lambda = 480 \pm 73$ ($\chi^2 = 10.3$, $P > 0.1$, 6 d.f.); a fit to a two-molecule model yields $\chi^2 = 81.7$, $P < 0.001$. **b**, Record of force and movement generated by a subspecies-c-coated bead captured by the optical trap. The trace shows movement of the bead along the microtubule. Displacement of the bead from the trap centre was measured over 100 s. The number of active subspecies-c molecules on the bead was 0.06 at $n = 30$ from **a**. **c**, Detail of movement of a

subspecies-c-coated bead with horizontal gridlines spaced at 8 nm. Multiple steps were observed. Arrowheads indicate backward movement of the bead ('slippage'). **d**, Pair-wise distance distribution of bead movement. Solid line, sum of seven gaussian components fitted to the data: peak position of the unit step was 8.3 nm (s.d. = 3.5 nm). Arrowheads indicate peak positions of each component. The compliance-attenuation factor 1.06 does not significantly affect results in **c** and **d**. **e**, Relationship between frequency of slippage and load on a subspecies-c-coated bead obtained from three traces. Error bars indicate s.d. All traces were filtered using a low-pass Butterworth filter (2-pole, -3dB at 100 Hz).

observed. The histogram of all pair-wise distances shows that movement occurred in integral multiples of 8-nm steps, the microtubule lattice repeat (Fig. 4d). At forces above 1 pN the bead often slipped backwards by one or two steps (arrowheads in Fig. 4c), the frequency of such slippages increasing as the load increased (Fig. 4e).

For processive motion, it is essential that the subspecies-c molecule remains bound to the microtubule throughout its movement, as when it detaches it quickly diffuses away. The head domain of subspecies-c is substantially larger than both kinesin and myosin and, given that it contains multiple ATP-binding sites, like other dyneins², the mechanism of its movement may be more complex. The negatively charged carboxy-terminal domain of the tubulins could interact electrostatically with positively charged dynein residues²⁵. This kind of weak interaction may be important in the processive movement of subspecies-c. The detachment rate of subspecies-c from microtubules is estimated to be $\sim 1/450$ per step (step length 8 nm divided by the length constant 3.6 μm), similar to that of kinesin (1/200–1/1,000 per step)⁹; backward slippage often occurs at high forces. In flagellar beating, the conversion of microtubules' sliding into rudimentary oscillatory bending waves occurs in small subsets of the axonemal structure³. The low detachment rate and strain-dependent slippage of subspecies-c may indicate that a structural component capable of limiting sliding resides in such dyneins. In addition, this strain-sensitive slippage could cause cooperative detachment of dynein molecules from a microtubule, leading to net backward microtubule movement until enough strain is released to allow those dyneins to re-engage. This cooperativity might be a source of oscillatory flagellar movement. Additional work is required to elucidate the unique structural and mechanical features of this single-headed dynein, including the relationship between processive movement and ATPase kinetics, that prevent detachment of the subspecies-c motor molecule from microtubules during both minus-end-directed movement and slippage under loaded conditions, but permit its high-velocity progression along the microtubule for many hundreds of steps. □

Methods

Protein preparation. We isolated inner-arm dynein subspecies-c from wild-type or outer-armless mutant (*oda1*) *Chlamydomonas reinhardtii* (strain 137c) as described¹⁵. Flagellar axonemes and high-salt extract of axonemes were obtained as described²⁶. Crude dynein extract was fractionated by high-performance liquid chromatography (HPLC) with a Mono Q HR5/5 anion-exchange column (Amersham-Pharmacia) by elution with a linear gradient from 120–500 mM KCl with 5 mM MgSO_4 , 1 mM EGTA, 1 mM dithiothreitol (DTT), 0.5 mM PMSF and 30 mM HEPES/KOH (pH 7.4). The pooled fraction of subspecies-c was purified by HPLC with a Mini Q PC3.2/3 column (Amersham-Pharmacia). We examined the purity and aggregation of subspecies-c by 3% SDS–PAGE with silver staining and by sucrose density-gradient centrifugation. No significant contamination of other subspecies and no aggregation were detected. Purified bovine brain tubulin²⁷ was polymerized into microtubules in assembly buffer (1 mM MgCl_2 , 1 mM EGTA, 1 mM GTP, 10% (v/v) dimethylsulphoxide, 80 mM PIPES/KOH, pH 6.9) at 37 °C, and they were stabilized by adding 10 μM taxol. Polarity-marked microtubules, with highest fluorescence at the minus ends, were prepared using NEM tubulin as described²¹. We determined subspecies-c and tubulin concentrations using Bradford assays.

Electron microscopy. For negative staining, we applied 5 μl of the sample solution directly to 400-mesh bare grids and stained it negatively with uranyl acetate as described²⁸. Accurate magnifications were calculated from the known microtubule repeats.

In vitro motility assays. *In vitro* assays were performed at 23 °C, basically as described¹⁵. A flow cell, volume 10 μl , was made from a glass slide and coverslip (18 mm $\times$ 18 mm), after cleaning with 0.1 M HCl and 70% ethanol, with two slivers of polycarbonate film 50 μm thick as spacers. We diluted fractionated subspecies-c (100 $\mu\text{g ml}^{-1}$) to various concentrations in buffer solution containing 30 mM HEPES/KOH (pH 7.4), 5 mM MgSO_4 , 1 mM DTT, 1 mM EGTA

and 1 mg ml^{-1} BSA. The flow cell was filled with 10 μl of the diluted subspecies-c solution, incubated for 5 min and washed with buffer solution, and then 20 μl of buffer solution containing 40 $\mu\text{g ml}^{-1}$ microtubules, 0.5 mM ATP, 1 mM DTT and 0.05% methylcellulose was infused into the flow cell. Microtubules were visualized on an Olympus microscope with dark-field illumination using a 100-W mercury light source. Images were recorded by a CCD camera (C2400-77, Hamamatsu Photonics), processed by contrast enhancement and brightness offset and stored on videotapes for subsequent computer analysis.

Single-molecule fluorescence technique for dynein surface density measurement. Cy3-EDA–ATP was prepared as described²⁹. The K_m of Cy3-EDA–ATP associated with subspecies-c-catalysed hydrolysis was determined by measuring the intensity of fluorescence of subspecies-c-coated glass surfaces at various concentrations of Cy3-EDA–ATP. Fluorescence intensity showed saturation kinetics with K_m of 52 nM. Cy3-EDA–ATP (1 nM) was then introduced into the flow cell and individual Cy3-EDA–AT(D)P molecules bound to subspecies-c on the surface were visualized by total internal reflection fluorescence microscopy^{17,18}. We recorded up to five fields every 10 s after 8-frame averaging and counted the fluorescent spots corresponding to Cy3-EDA–nucleotide bound to subspecies-c. We estimated the number of subspecies-c molecules on the surface by multiplying the number of spots observed by 53 ($\{K_m + [\text{Cy3-EDA-ATP}]\}/[\text{Cy3-EDA-ATP}]$).

Optical trap nanometry. Polystyrene carboxylated beads, 1 μm in diameter, were preincubated in buffer (100 mM NaCl, 10 mM Tris, pH 10) supplemented with 1 mg ml^{-1} casein. Beads were then incubated in assay buffer (5 mM MgSO_4 , 1 mM EGTA, 30 mM HEPES/KOH, pH 7.4) for 10 min with varying concentrations of subspecies-c. Coated beads were added to taxol-stabilized rhodamine-labelled microtubules (10 $\mu\text{g ml}^{-1}$ taxol) adsorbed onto a coverslip. The experimental set-up for the optical trap (Nd:YAG laser power 500 mW at 1,064 nm, Spectra Physics) was similar that of ref. 23. We measured displacement of the beads using a quadrant photodiode system²³ with 0.1 nm spatial (noise-equivalent value) and 0.02 ms temporal resolution. Trap stiffness (0.017 pN nm^{-1}) and bead-to-dynein linkage stiffness (0.1–0.3 pN nm^{-1}) were determined from the variances of the thermal fluctuations of beads. Displacement of beads within the trap was converted to displacement of subspecies-c by multiplying by an attenuation factor, the ratio of the sum of bead-to-dynein linkage stiffness and trap stiffness to bead-to-dynein linkage stiffness²³. To determine slippage frequency, displacements of individual beads from the trap centre were divided into 20-nm bins, and in each bin we counted backward bead movements of more than 4 nm that were followed by a dwell phase of more than 40 ms. These backward movement counts were divided by the total time during which the bead remained in each displacement bin.

Received 12 April; accepted 16 June 1999.

- Gibbons, I. R. Cilia and flagella of eukaryotes. *J. Cell Biol.* **91**, 107s–124s (1981).
- Holzbaur, E. L. F. & Vallee, R. B. Dyneins: Molecular structure and cellular function. *Annu. Rev. Cell Biol.* **10**, 339–372 (1994).
- Kamiya, R. Exploring the function of inner and outer dynein arms with *Chlamydomonas* mutants. *Cell Motil. Cytoskeleton* **32**, 98–102 (1995).
- Brokaw, C. J. & Kamiya, R. Bending patterns of *Chlamydomonas* flagella: IV. Mutants with defects in inner and outer dynein arms indicate differences in dynein arm function. *Cell Motil. Cytoskeleton* **8**, 68–75 (1987).
- Takahashi, K. & Kamimura, S. Dynamic aspects of microtubule sliding in sperm flagella. *J. Submicrosc. Cytol.* **15**, 1–3 (1983).
- Oiwa, K. & Takahashi, K. The force-velocity relationship for microtubule sliding in demembrated sperm flagella of the sea urchin. *Cell Struct. Funct.* **13**, 193–205 (1988).
- Kamimura, S. & Kamiya, R. High-frequency nanometre-scale vibration in 'quiescent' flagellar axonemes. *Nature* **340**, 476–478 (1989).
- Shingyoji, C. *et al.* Dynein arms are oscillating force generators. *Nature* **393**, 711–714 (1998).
- Block, S. M., Goldstein, L. S. B. & Schnapp, B. J. Bead movement by single kinesin molecules studied with optical tweezers. *Nature* **348**, 348–352 (1990).
- Uyeda, T. Q. P., Warrick, H. M., Kron, S. J. & Spudis, J. A. Quantized velocities at low myosin densities in an *in vitro* motility assay. *Nature* **352**, 307–311 (1991).
- Goodenough, U. W. *et al.* High-pressure liquid chromatography fractionation of *Chlamydomonas* dynein extracts and characterization of inner-arm dynein subunits. *J. Mol. Biol.* **194**, 481–494 (1987).
- Goodenough, U. & Heuser, J. Structural comparison of purified dynein proteins with *in situ* dynein arms. *J. Mol. Biol.* **180**, 1083–1118 (1984).
- Amos, L. A. Brain dynein crossbridges microtubules into bundles. *J. Cell Sci.* **93**, 19–28 (1989).
- Samsó, M., Radermacher, M., Frank, J. & Koonce, M. P. Structural characterization of a dynein motor domain. *J. Mol. Biol.* **276**, 927–937 (1998).
- Kagami, O. & Kamiya, R. Translocation and rotation of microtubules caused by multiple species of *Chlamydomonas* inner-arm dynein. *J. Cell Sci.* **103**, 653–664 (1992).
- Gee, M. A., Heuser, J. E. & Vallee, R. B. An extended microtubule-binding structure within the dynein motor domain. *Nature* **390**, 636–639 (1997).
- Funatsu, T. *et al.* Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature* **374**, 555–559 (1995).
- Oiwa, K. *et al.* Microscopic kinetic measurements of single Cy3-EDA-ADP molecules interacting with myosin filaments *in vitro*. *Biophys. J.* **72**, A180 (1997).

19. Howard, J., Hudspeth, A. J. & Vale, R. D. Movement of microtubules by single kinesin molecules. *Nature* **342**, 154–158 (1989).
20. Hancock, W. O. & Howard, J. Processivity of the motor protein kinesin requires two heads. *J. Cell Biol.* **140**, 1395–1405 (1998).
21. Howard, J. & Hyman, A. A. Preparation of marked microtubules for the assay of the polarity of microtubule-based motors by fluorescence microscopy. *Methods Cell Biol.* **39**, 105–113 (1993).
22. Cole, D. G. et al. *Chlamydomonas* kinesin-II-dependent intraflagellar transport (IFT): IFT particles contain proteins required for ciliary assembly in *Caenorhabditis elegans* sensory neurons. *J. Cell Biol.* **141**, 993–1008 (1998).
23. Kojima, H., Muto, E., Higuchi, H. & Yanagida, T. Mechanics of single kinesin molecules measured by optical trapping nanometry. *Biophys. J.* **73**, 2012–2022 (1997).
24. Svoboda, K. & Block, S. M. Force and velocity measured for single kinesin molecules. *Cell* **77**, 773–784 (1994).
25. Paschal, B. M., Obar, R. A. & Vallee, R. B. Interaction of brain cytoplasmic dynein and MAP2 with a common sequence at the C terminus of tubulin. *Nature* **342**, 569–572 (1989).
26. King, S. M., Otter, T. & Witman, G. B. Purification and characterization of *Chlamydomonas* flagellar dyneins. *Methods Enzymol.* **134**, 291–306 (1986).
27. Vallee, R. B. Reversible assembly purification of microtubules without assembly-promoting agents and further purification of tubulin, microtubule-associated proteins, and MAP fragments. *Methods Enzymol.* **134**, 89–104 (1986).
28. Katayama, E. The effects of various nucleotides on the structure of actin-attached myosin subfragment-1 studied by quick-freeze deep-etch electron microscopy. *J. Biochem.* **106**, 751–770 (1989).
29. Jameson, D. M. & Eccleston, J. F. Fluorescent nucleotide analogs: Synthesis and applications. *Methods Enzymol.* **278**, 363–390 (1997).

Acknowledgements. We thank M. Anson for discussion and comments on the manuscript and M. Kikumoto for single molecule measurements. This work was partly supported by the Hyogo Science and Technology Foundation (K.O. and H.K.) and a grant-in-aid from the Ministry of Education, Science and Culture of Japan (K.O.).

Correspondence and requests for materials should be addressed to K.O. (e-mail: oiwa@crl.go.jp).

Myosin-V is a processive actin-based motor

Amit D. Mehta*, Ronald S. Rock*, Matthias Rief*, James A. Spudich*, Mark S. Mooseker† & Richard E. Cheney‡

* Department of Biochemistry, Stanford University Medical Center, Stanford, California 94305, USA

† Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA

‡ Department of Cell and Molecular Physiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA

Class-V myosins, one of 15 known classes of actin-based molecular motors, have been implicated in several forms of organelle transport^{1–3}, perhaps working with microtubule-based motors such as kinesin^{2–4,6}. Such movements may require a motor with mechanochemical properties distinct from those of myosin-II, which operates in large ensembles to drive high-speed motility as in muscle contraction⁷. Based on its function and biochemistry, it has been suggested that myosin-V may be a processive motor^{7,8} like kinesin^{9,10}. Processivity means that the motor undergoes multiple catalytic cycles and coupled mechanical advances for each diffusional encounter with its track. This allows single motors to support movement of an organelle along its track. Here we provide direct evidence that myosin-V is indeed a processive actin-based motor that can move in large steps approximating the 36-nm pseudo-repeat of the actin filament.

Myosin-V purified from chick brain consists of two heavy chains, each with an amino-terminal head (motor) domain, a neck region

supported by six light chains, a tail with a proximal coiled coil and a distal globular domain presumed to bind cargo and/or specify subcellular localization^{1,11,12}. Myosin-V exhibits actin-activated ATP turnover and moves actin filaments *in vitro*¹¹. To examine the dependence of movement on the number of motors, we measured this *in vitro* motility in 2 mM ATP for a wide range of myosin-V surface densities. At 54 molecules per μm^2 on the surface (600 pM added to flow chamber), myosin-V supported continuous and smooth actin-filament movement ($287 \pm 22 \text{ nm s}^{-1}$, $n = 60$). We continued to observe motility with myosin-V concentrations as low as 2.7 molecules per μm^2 ($311 \pm 70 \text{ nm s}^{-1}$, $n = 20$). At these very low surface densities of myosin-V, many actin filaments appeared to be tethered to the surface by a single contact point as they moved (Fig. 1a). Moreover, the moving filaments swivelled about the connection point, similar to behaviour reported for single kinesin molecule attachments⁹. Actin velocity did not fall as the myosin-V surface density was decreased from 1,000 to 2.7 molecules per μm^2 (data not shown).

To determine whether such motility is driven by single molecules or by coincident co-localization of several non-processive motors, we measured actin filament landing rates and distances moved over several decades of myosin-V surface density. We observed first-power dependence of the landing rate on surface density (Fig. 1b), as expected for a processive motor^{9,13}. There was a gradual transition, from zero to one, in the fraction of filaments moving more than their length before releasing from the surface, as the myosin-V density on the surface was increased (Fig. 1c). The data (Fig. 1c) could be fitted to a functional form describing movement driven by a single molecule⁹: $1 - (\rho/\rho_0)\exp(-\rho/\rho_0)/(1 - \exp(-\rho/\rho_0))$, where ρ is surface density and ρ_0 is a fit parameter. These signatures of processivity¹³, shared by kinesin^{9,13} and myosin-V, differ markedly from those observed with single-headed kinesins¹³, where four to six molecules are believed to be necessary and sufficient to propel sustained microtubule movement.

To examine single-molecule movement at high resolution, we used an optical trapping assay^{14,15}, with 10 pM myosin-V loaded in the flow cell (0.9 molecules per μm^2 on the surface). From system geometry, we estimate that we sample $0.09 \mu\text{m}^2$ along the spherical platform surface, leading to an estimate of 0.08 accessible molecules per surface platform tested. As expected, we observed no binding events on ~ 90 –95% of the surface platforms sampled. Were more than one molecule required for the behaviour observed, we would have detected no binding events in 99.7% of the platforms sampled¹⁰. In most binding events detected, the actin filament was bound and pulled for three-to-five step increments before dwelling at a stall point, detaching and returning to the baseline (Fig. 2a–c). Myosin-V stepped to a stall point, with stall force $3.0 \pm 0.3 \text{ pN}$ ($n = 25$) at 2 mM ATP, considerably less than the 5–8 pN observed for kinesin^{16–19}. A lower stall force is expected for a motor with a larger step size (see below) if the maximum work output per step is similar for the two motors.

Load-dependent arrest of myosin-V movement may proceed through either or both of two mechanisms: first, load reduces the ATPase cycling rate by affecting a load-dependent biochemical transition (for example, release of the nucleotide hydrolysis product), as suggested for RNAP²⁰; or, second, load affects the prob-

Table 1 Summary of step-size measurements

[ATP]	Frequency	Forward			Backward		
		<i>n</i>	Mean (nm)	s.d.	<i>n</i>	Mean (nm)	s.d.
2 mM	20 Hz	98	35.0	8.9	16	–36.3	8.5
2 mM	20 Hz	74	34.3	11.2	27	–37.0	13.2
2 mM	10 Hz	71	35.6	8.3	9	–38.8	9.2
2 mM	100 Hz	152	30.3	5.3	9	–27.5	11.8
10 μM	20 Hz	19	37.0	10.3	7	–29.0	10.2
1 μM	20 Hz	68	38.0	6.6	39	–36.0	9.1

All steps are measured for the left bead in Fig. 3a. Examples of forward events are as shown in Fig. 3b. Occasional backward events (not shown in Fig. 3) were seen in this experiment, similar to those seen in Fig. 2e.

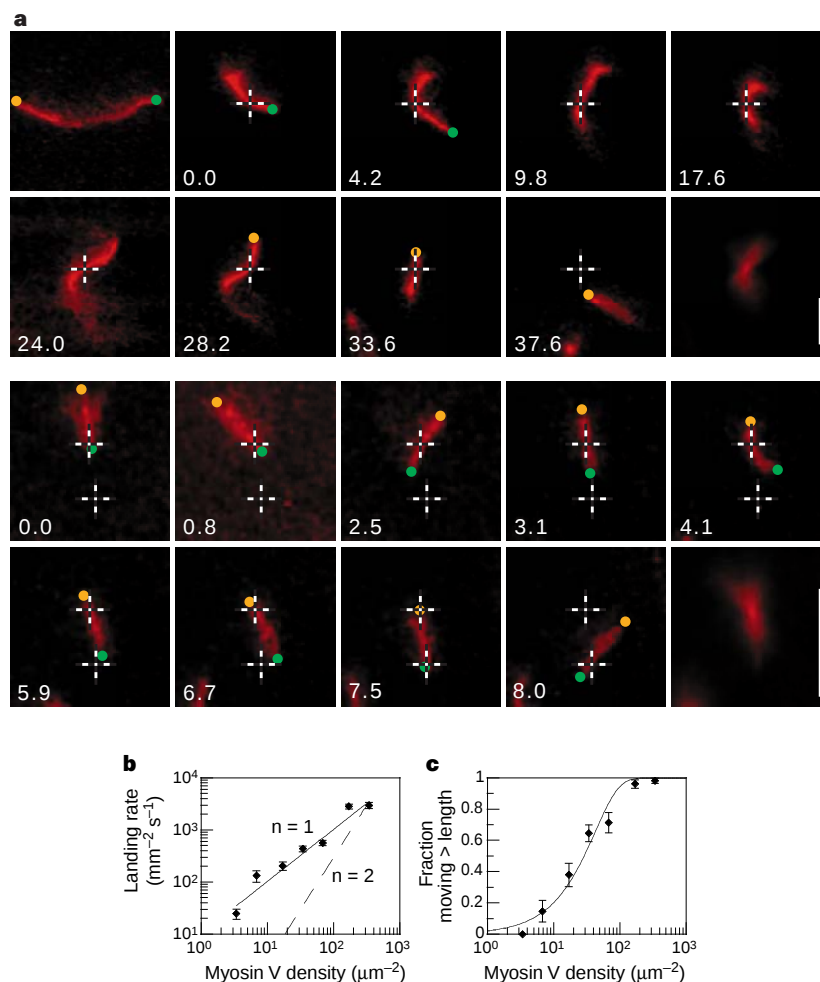


Figure 1 Actin-filament gliding. **a**, Actin filament moving over surface coated with 3.6 myosin-V molecules per μm^2 (2 mM ATP). The leading (green dots) and trailing (yellow dots) ends of the filament are marked when within the focal plane. Panels 1–10 (left right): time course (s) of a 16- μm actin filament before (panel 1) and after (panels 2–9) it bound to the surface. The filament remained connected at the in-focus contact point (crosshair), while the ends diffused in and out of the focal plane. Panel 10 shows the average image throughout the entire range of

movement. Panels 11–20: movement (at 5.4 myosin-V molecules per μm^2) of a shorter filament whose diffusive motion was confined to the focal plane. The second crosshair points to a second attachment point. Scale: 5 μm . **b**, Actin-filament landing rates ($\text{mm}^{-2}\text{s}^{-1}$) as a function of surface myosin-V density. **c**, Fraction of filaments that moved more than their length before dissociating, as a function of surface density.

ability and direction of the mechanical step resulting from a catalytic cycle, as suggested for kinesin^{16,17}. Our data favour a combination of the two for myosin-V. In support of the first proposal, as the load rises, dwell periods preceding forward steps become longer at 2 mM ATP but not at 1 μM ATP (Fig. 2d). This could mean, for example, that a load-decelerated ADP release no longer affects the ATPase cycling rate at 1 μM ATP, as ATP binding has become rate-limiting. Moreover, the load dependence of dwell periods at 2 mM ATP, rising sharply near stall, is consistent with those suggested to explain load-dependent chemistry (Fig. 2d)²⁰. Consistent with the second mechanism, however, and as observed with kinesin^{16–19}, we frequently observed backward stepping at high load. Unlike anything observed with kinesin, we also observed sequences of two or three consecutive backward steps (Fig. 2e).

We performed optical trap experiments at 5 mM, 2 mM, 10 μM and 1 μM ATP. As the ATP concentration was dropped, dwell durations between steps at low load (where load and durations are uncorrelated) grew longer (0.08 ± 0.01 ($n = 37$), 0.10 ± 0.02 ($n = 64$), 0.28 ± 0.08 ($n = 12$) and 1.56 ± 0.12 ($n = 212$) s, respectively). Dwell-period durations at 1 μM ATP obey single-exponential statistics, indicating^{19,21,22} that one ATP is coupled to one mechanical advance. Multiplying the dwell times by gliding velocities (54 molecules per μm^2 , $n = 60$ at each ATP) at these four

ATP concentrations yields 22–30 nm, 23–35 nm, 20–36 nm and 21–35 nm, respectively (range from propagated uncertainty). These remarkably consistent numbers overlap substantially the range of direct step measurements, to which we now turn.

The step sizes observed in the standard optical trap assay (Fig. 2a–c) are limited by compliance in the various connections (bead to actin filament and myosin-V to surface), and perhaps in the molecule itself²³. These compliances cause the advance in bead position to reflect only part of myosin-V's advance along the actin filament. To circumvent compliance issues, we placed large-amplitude sinusoidal or triangle wave oscillations on one of the two trapped beads (Fig. 3a, left bead). When the bead in the oscillated trap was pulled by myosin-V from its trap centre, we observed clipping in the bead oscillation, presumably because the system, including the myosin-V molecule and its surface attachment, was pulled taut. By examining the clipped level, we measured the tether length between the trapped bead and the myosin-V molecule, which decreased in increments as myosin-V stepped along the actin filament (Fig. 3a, b). Six different data sets each gave a broadly distributed step-value distribution, five centred about 34–38 nm and one about 30 nm (Fig. 3c, Table 1).

A step size of 30–38 nm (exceeding the 4–17-nm range of myosin-II step estimates²⁴ and the 8-nm kinesin step estimate²⁵) is

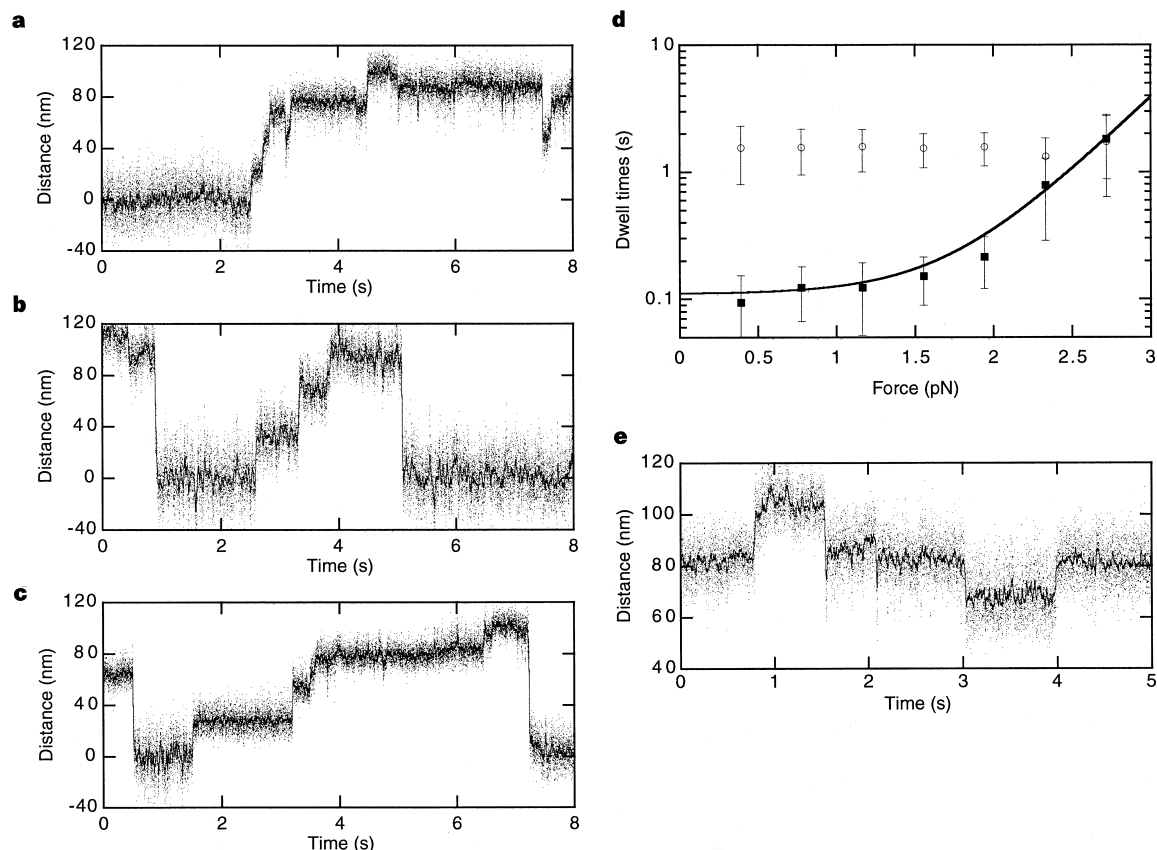


Figure 2 Processive stepping by myosin-V observed using the dual-beam optical trap. **a–c**, Sample traces of stepping behaviour at 2 mM, 10 μM and 1 μM ATP, respectively. Dwell periods T_i at 1 μM ATP fit single exponential statistics, as judged by $(\text{mean}\{T_i | T_i > \tau_o\} - \tau_o)$ remaining constant against varying τ_o (ref. 30).

d, Mean dwell time as a function of load, at 2 mM ATP (squares) and at 1 μM ATP (circles). **e**, Backward stepping behaviour (at 1.5 s and 3 s) at high load in 1 μM ATP.

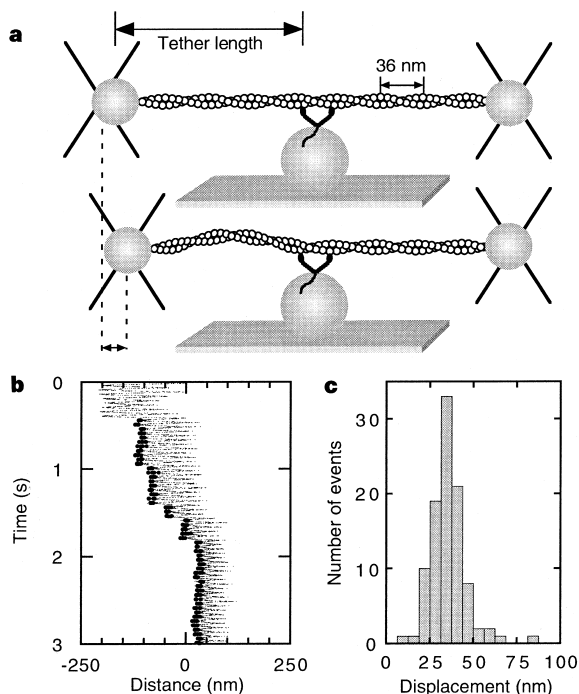


Figure 3 Step measurements. **a**, The scheme of the oscillation experiments (see Methods). **b**, Sample trace at 2 mM ATP. The myosin molecule bound the actin just before the 0.5-s mark. After this, the bead ceased to follow the driving triangle wave for a fraction of each oscillation. **c**, Sample histogram of step increments so measured.

plausible given that $\sim 100^\circ$ angle between two ~ 23 nm necks can subtend this distance⁷. Interestingly, this step size may allow the molecule to step along the 36-nm pseudo-repeat of the actin helix. This would allow a single myosin-V molecule and its cargo to move along a single filament without spiralling around it. However, the spread in the step-size measurements (Table 1) exceeds the $< \sim 4$ -nm measurement precision. This spread raises the possibility that myosin-V may not move precisely along the pseudo-repeat, but rather can bind along the filament at different points while still approximating the pseudo-repeat on average. Alternatively, a randomizing effect in the assay²⁶ (for example, limited actin rotation, shifting randomly the next myosin binding site on the actin filament) may underlie the distribution width.

These data allow us to estimate the average minimum number of steps a single myosin-V molecule can make at zero load before releasing the actin for long enough to diffuse away. We estimate a minimum mean Λ of travel lengths L_i of 1.6 μm ($n = 27$) (measured from surface attachment to release in Fig. 1 assays, 2 mM ATP, 3.5 molecules per μm², with nodal swivelling). We compute this using $(\text{mean}\{L_i | L_i > L_c\} - L_c)$, thus assuming that it is exponentially distributed¹⁰ with a low-end reliable measurement threshold $t_i = 2$ μm. As many filaments were moved by myosin-V to their ends and released only then, Λ is certainly an underestimate. From this we estimate that myosin-V can take at least 40–50 steps on average before releasing actin.

Although class-V myosins are related in sequence and structure to muscle myosin-II, they face a significantly different biological task: movement of cargo by relatively few motors in regions containing relatively few actin filaments. A processive motor could more easily perform the roles suggested by myosin-V-dependent endoplasmic

reticulum vesicle transport², messenger RNA transport in yeast²⁷ and organelle transport generally^{1,3–5}. Moreover, prolonged contact between myosin-V and its actin track may underlie its proposed role as an organelle tether. Such a role has been indicated by myosin-V-dependent retention of smooth endoplasmic reticulum with dendritic spines of Purkinje neurons⁴ and retention of melanosomes within the dendrites of melanocytes^{1,3}. The findings reported here, using single-molecule analysis, fit well with the expected roles of class-V myosins and demonstrate the existence of processivity within the myosin superfamily. □

Methods

Gliding filament assay. In all assays described, motility buffers¹⁵ (23 °C) included 25 mM imidazole HCl, pH 7.4, 25 mM KCl, 1 mM EGTA, 10 mM DTT, with variable Mg-ATP, an ATP-regenerating system (1 mM phosphocreatine, 0.1 mg ml⁻¹ creatine phosphokinase) and an oxygen-scavenging system to retard photobleaching (25 µg ml⁻¹ glucose oxidase, 45 µg ml⁻¹ catalase and 1% glucose). Myosin-V was purified as described²⁸. To quantify filament-landing rates (Fig. 1b), we scored as 'landed' a filament that had touched down and moved for >0.5 µm over 2 s. Landing assays were performed at 2 mM ATP with constant actin concentration (5 µg ml⁻¹). All surface myosin-V densities given assume that every molecule entered the flow cell, none suffered damage, and half adsorbed to each surface of the flow cell. Electron microscopy (EM) analysis¹¹ and atomic-force microscopy imaging in solution (not shown) indicated that myosin-V molecules adsorbed at 100 nM to a surface do not oligomerize, and analytical ultracentrifugation in the storage buffer showed a single well-defined boundary with an *s* value of 12.52 (O. C. Rodriguez and R. E. Cheney, personal communication). We observed the same actin landing rates at 6.8 myosin-V molecules per µm² after storage, dilution and adsorption of myosin-V in the motility assay buffer¹⁵, myosin-V storage buffer²⁸ or the buffer used in the EM experiments¹¹. Kinetic arguments exclude aggregation of protein on the surface⁹. Working at low densities required pre-adsorbing 10 µg ml⁻¹ BSA on the surface (2 min incubation), before the addition of myosin. Similar pre-adsorption was necessary to observe motility under very low densities of kinesin^{9,10}. To measure filament speed, we recorded and digitized visual fields at 30 frames s⁻¹ and identified filament end positions. At low density, when the filament ends diffused into the image focal plane, we measured the contour length separating the fixed contact point to the filament end. As position (over 2–4 µm in 5–30 s) advances in time with constant position-detection error (0.2 µm), we can determine velocities reliably by measuring over sufficient length intervals. The error bars in Fig. 1b, c, are standard errors derived from counting statistics¹⁵. Velocity measurements are reported with standard deviations.

Optical trap. A single actin filament attached at both ends to optically trapped 1-µm-diameter polystyrene beads was stretched to tension and moved near to surface-bound silica spheres, decorated sparsely with myosin-V molecules. Both polystyrene beads were tracked with nanometre and millisecond resolution¹⁵. We used a biotin-avidin system to link actin to the polystyrene beads²⁹.

Load-dependent dwell times. To compute the load burdening a molecule, we summed the forces exerted on the molecule by each of the two traps (16 and 30 fN nm⁻¹). We measured the duration of dwell periods separating step transitions and preceding forward-directed ones. Uncertainties in these and other dwell times are computed as described³⁰. Assuming load-dependent chemistry at 2 mM ATP, we fitted the data with $\tau = \tau_1 + \tau_2 \exp(F\delta/k_B T)$ ²⁰, where τ is the dwell time, τ_1 represents load-independent transitions, τ_2 load-dependent transitions, F the load, δ the 'characteristic distance' over which load affects stall (not necessarily the step size) and $k_B T$ the thermal energy, to generate the solid curve shown in Fig. 2d. We extracted the following best-fit parameters: $\tau_1/\tau_2 > 100$, indicating that load-independent biochemical transitions are >100-fold slower than mechanical ones at vanishing load; and δ ranges from 10 to 15 nm, on the order of but less than the measured step size. These conclusions are unaffected by changes in binning parameters.

Step-size measurements (see Fig. 3). A 270-nm peak-to-peak triangle wave (double arrow separating dashed lines in Fig. 3a) was applied to the optical trap on the left (stiffness 16 fN nm⁻¹), moving the actin filament between a taut state and a compliant state. As myosin binds and pulls the filament to the right, the total trapped bead-to-surface platform linkage oscillates between states of very high and very low tension. To measure step sizes we monitored the mean position of the clipped (artificially darkened in Fig. 3b) part in each oscillation cycle as the bead continued to step. We measured steps (in nm) with oscillations at 10, 20 and 100 Hz at 2 mM ATP, and 20 Hz at 10 µM and 1 µM ATP (Table 1). Step amplitudes were only weakly correlated with displacement from baseline, indicating that all compliance had been similarly removed throughout the range of stepping. We scored only transitions with more than three cycles of dwell at a consistent position on both starting and ending level. Outliers may reflect the occasional incidence of two steps within one oscillation cycle. □

Received 23 March; accepted 23 June 1999.

- Wu, X., Bowers, B., Rao, K., Wei, Q. & Hammer, J. A. Visualization of melanosome dynamics within wild-type and dilute melanocytes suggests a paradigm for myosin-V function *in vivo*. *J. Cell Biol.* **143**, 1899–1918 (1998).
- Tabb, J. S., Molyneux, B. J., Cohen, D. L., Kuznetsov, S. A. & Langford, G. M. Transport of ER vesicles on actin filaments in neurons by myosin-V. *J. Cell Sci.* **111**, 3221–3234 (1998).
- Rogers, S. L. & Gelfand, V. I. Myosin cooperates with microtubule motors during organelle transport in melanophores. *Curr. Biol.* **8**, 161–164 (1998).
- Mermall, V., Post, P. L. & Mooseker, M. S. Unconventional myosins in cell movement, membrane traffic and signal transduction. *Science* **279**, 527–533 (1998).
- Evans, L. L., Lee, A. J., Bridgman, P. C. & Mooseker, M. S. Vesicle-associated brain myosin-V can be activated to catalyze actin-based transport. *J. Cell Sci.* **111**, 2055–2066 (1998).
- Huang, J. D. *et al.* Direct interaction of microtubule- and actin-based transport motors. *Nature* **397**, 267–270 (1999).
- Howard, J. Molecular motors: structural adaptations to cellular functions. *Nature* **389**, 561–567 (1997).
- Nascimento, A. A. C., Cheney, R. E., Tauhata, S. B. F., Larson, R. E. & Mooseker, M. S. Enzymatic characterization and functional domain mapping of brain myosin-V. *J. Biol. Chem.* **271**, 17561–17569 (1996).
- Howard, J., Hudspeth, A. J. & Vale, R. D. Movement of microtubules by single kinesin molecules. *Nature* **342**, 154–158 (1989).
- Block, S. M., Goldstein, L. S. & Schnapp, B. J. Bead movement by single kinesin molecules studied with optical tweezers. *Nature* **348**, 348–352 (1990).
- Cheney, R. E. *et al.* Brain myosin-V is a two-headed unconventional myosin with motor activity. *Cell* **75**, 13–23 (1993).
- Reck-Peterson, S. L., Novick, P. J. & Mooseker, M. S. The tail of a yeast class V myosin, myo2p, functions as a localization domain. *Mol. Biol. Cell* **10**, 1001–1017 (1999).
- Hancock, W. O. & Howard, J. Processivity of the motor protein kinesin requires two heads. *J. Cell Biol.* **140**, 1395–1405 (1998).
- Finer, J. T., Simmons, R. M. & Spudich, J. A. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature* **368**, 113–119 (1994).
- Mehta, A. D., Finer, J. T. & Spudich, J. A. Use of optical traps in single-molecule study of nonprocessive biological motors. *Methods Enzymol.* **298**, 436–459 (1998).
- Svoboda, K. & Block, S. M. Force and velocity measured for single kinesin molecules. *Cell* **77**, 773–784 (1994).
- Meyhofer, E. & Howard, J. The force generated by a single kinesin molecule against an elastic load. *Proc. Natl Acad. Sci. USA* **92**, 574–578 (1995).
- Coppin, C. M., Pierce, D. W., Hsu, L. & Vale, R. D. The load dependence of kinesin's mechanical cycle. *Proc. Natl Acad. Sci. USA* **94**, 8539–8544 (1997).
- Kojima, H., Muto, E., Higuchi, H. & Yanagida, T. Mechanics of single kinesin molecules measured by optical trapping nanometry. *Biophys. J.* **73**, 2012–2022 (1997).
- Wang, M. D. *et al.* Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902–907 (1998).
- Schnitzer, M. J. & Block, S. M. Kinesin hydrolyses one ATP per 8-nm step. *Nature* **388**, 386–390 (1997).
- Hua, W., Young, E. C., Fleming, M. L. & Gelles, J. Coupling of kinesin steps to ATP hydrolysis. *Nature* **388**, 390–393 (1997).
- Howard, J. & Spudich, J. A. Is the lever arm of myosin a molecular elastic element? *Proc. Natl Acad. Sci. USA* **93**, 4462–4464 (1996).
- Mehta, A. D. & Spudich, J. A. Single myosin molecule mechanics. *Adv. Struct. Biol.* **5**, 229–270 (1999).
- Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
- Molloy, J. E., Burns, J. E., Kendrick-Jones, J., Tregear, R. T. & White, D. C. Movement and force produced by a single myosin head. *Nature* **378**, 209–212 (1995).
- Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).
- Cheney, R. E. Purification and assay of myosin-V. *Methods Enzymol.* **298**, 3–18 (1998).
- Ishijima, A. *et al.* Simultaneous observation of individual ATPase and mechanical events by a single myosin molecule during interaction with actin. *Cell* **92**, 161–171 (1998).
- Veigel, C. *et al.* The motor protein myosin-I produces its working stroke in two steps. *Nature* **398**, 530–533 (1999).

Acknowledgements. We thank O. Rodriguez, C. Pennisi and C. Pearson for assistance in the purification of myosin-V; T. Sulchek for assistance in AFM imaging; J. Sellers for helpful discussions; and L. Evans, V. Mermall, H. Warrick, and M. Heidecker for their comments on the manuscript.

Correspondence and requests for materials should be addressed to A.D.M. or R.E.C.

The two cultures

Cell and tissue culture are the focus in this week's New on the Market, including serum-free media and see-through culture plates.



Reservoir dodge: HERAccl maintains humidity.

HERAccl incubators

From Kendro www.kendro.com
CO₂ incubators with automatic disinfection

The ContraCon 90°C automatic disinfection routine sterilizes the entire inner chamber in the HERAccl CO₂ incubator. The aim is to protect against long-term contamination. Suitable for a broad range of cell and tissue culture applications, the 150-litre capacity HERAccl features a smooth inner chamber with rounded corners on all sides for easy cleaning. The HERAccl's three-litre water reservoir is also disinfected automatically. This reservoir is a humidification device, eliminating the need for the traditional water tray.

Reader Service No. 100

Cell Proliferation Assay

From Genzyme www.genzyme.com
Tracking cell proliferation and viability

This assay from Genzyme Diagnostics measures cell proliferation and viability in response to factors such as cytokines, growth factors and experimental pharmaceuticals. The MTT Cell Proliferation & Viability Assay is a nonisotopic *in vitro* assay that uses the 96-well format. The assay is based on reduction of the tetrazolium salt 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-tetrazolium bromide (MTT) to an insoluble formazan dye by cellular enzymes associated with metabolic activity. The rate of tetrazolium reduction is proportional to the rate of cell proliferation. The assay may be used in systems with co-factors and with cell culture media that contain reducing agents.

Reader service No. 101

Cellmate

From The Automation Partnership

www.autoprt.co.uk

Automating tissue culture

Cellmate automates conventional cell culture processes in roller bottles and T-flasks, allowing scale-up without process change. The latest version of Cellmate has more sophisticated software and a new robot arm to make program changes easier. Cell harvesting by a new scraper blade protects delicate cell-surface proteins from damage. Cellmates are used in the production of viral vaccines and therapeutic proteins, as well as for high-throughput screening assays.

Reader Service No. 102

cellgro Complete

From Mediatech

www.cellgro.com

Serum-free media for both attached and suspended mammalian cell lines

'cellgro Complete' was originally formulated for RPMI, DMEM, DMEM/F12 and McCoy's 5A basal media users, but can also be adapted for many cell-culture protocols currently using other basal media. It has a low protein concentration and contains no insulin, transferrin, hormones or other growth factors. The low protein level and lack of undefined components minimizes interference with downstream processing.

Reader Service No. 103

StemSpan SFEM

From Stem Cell Technologies

www.stemcell.com

Flexible serum-free media

StemSpan SFEM (serum-free expansion medium) is optimized for haematopoietic cells. With appropriate cytokines, it can be used to culture both human and murine haematopoietic cells. StemSpan SFEM also supports the generation of dendritic cells from CD14⁺ and CD34⁺ progenitor cell populations. From the bottle (100 and 500 ml) StemSpan SFEM contains no cytokines, making it possible to tailor the medium for specific cells by adding cytokines separately.

Reader Service No. 104

Nutri-Mister

From Bellco

www.bellcoglass.com

Get to the problem of the roots

Nutri-Mister is an ultrasonic generator that provides a nutrient-laden, oxygen-rich mist to plant roots or to plant or other tissue. For



Feed those plants: Nutri-Mister at rest.

whole plant culture, roots are suspended in an enclosed plant holder/root chamber. Plants are supported by clay pebbles in slatted plastic cups, which allow the roots to protrude into the root chamber. For tissue culture, sterile plant tissues are placed on a perforated tray or screen, which sits inside a transparent, polycarbonate culture chamber.

Reader Service No. 105

Nephrogen

From Celox

celox.bioreagent.com

Serum-free growth medium for '293' kidney cells

Nephrogen is a serum-free growth medium developed specifically for the culturing of human embryonic kidney (293) cells. This low-protein medium contains no bovine derived proteins. As part of the Nephrogen system, Celox also produces a non-enzymatic dissociation solution to be used instead of trypsin.

Reader Service No. 106



Nephrogen: for 293 reasons.

HEK-293 media

From Sigma

www.sigma-aldrich.com

Another growth medium for '293' kidney cells

This new line of serum-free, low-protein media for the growth of human embryonic kidney 293 (HEK293) cells is recommended for human adenovirus production, recombinant protein production, and vaccine production. Testing of Sigma's serum-free HEK-293 culture media indicates faster cell growth when compared to DME:F12 media supplemented with 5% fetal bovine serum.

Reader Service No. 107

Suprafusion 2500

From Brandel

www.brandel.com

Automation in receptor binding assays

Suprafusion 2500 Systems offer automatic, programmable control, perfusing tissues with up to 9 reagents each and collecting up to 20 effluents per sample. The 6, 12 and 20-channel systems collect into standard 5 or 7 ml vials in a generic collection tray. The 18-channel system is designed for use with specific liquid scintillation counters as effluents are collected into vials preloaded into the counter's rack. Models are available for

use with Beckman, LKB and Packard counters. All models can be supplied for perfusing larger tissue samples as well as performing other tests in wellplates.

Reader Service No. 108

Custom cultures

From TCS Biologicals

Normal human cells and custom cell cultures

In addition to normal human cells and media, TCS Biologicals offers custom cell culture for functional genomic and proteomic studies, and Liverbeads, which are encapsulated hepatocytes for metabolism analysis allowing users to perform functional tests without needing to isolate cells. Services include growing cells and treating them with cytokines or growth factors. Customers can be supplied with cell pellets, lysates, nucleic acid or other material as required.

Reader Service No. 109

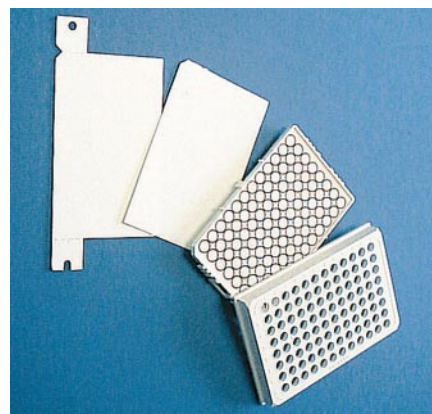
EasyTab-C filtermats

From Packard

www.packardinst.com

Simplify cell proliferation assays

The EasyTab-C is designed for use with Packard's FilterMate harvester and the



Harvest home: Packard's EasyTab-C

reusable OmniFilter filter holder. The filter is self-aligning in the harvester, using two locator holes, which mate to the alignment pins of the harvester. After cells are harvested, EasyTab-C fits directly into the OmniFilter holders once alignment tabs are removed. After scintillation cocktail is dispensed on the filtermat, the OmniFilter holder is snapped together and sealed for counting in the TopCount Microplate Scintillation and Luminescence Counter.

Reader Service No. 110

ADVERTISEMENTS

REVOLUTION IN THE VIAL?

All facets of the biotechnology in the BIOTECHNICA.

BIOTECHNICA starts discussion between commerce and science. With trade fair, innovation round-table, congress and Financial World.

Detailed information under
www.biotechnica.de

BIO BIOTECHNICA
INTERNATIONAL TRADE FAIR
FOR BIOTECHNOLOGY
HANNOVER 5.-7. OCTOBER 1999

Deutsche Messe AG · Messengelände · D-30521 Hannover, Germany
Phone: (+49) 511/89-0 · Fax: (+49) 511/89-326 26

READER ENQUIRY NO. 68

GMD Secretion Assay

From One Cell Systems www.onecell.com

Screen cell lines for rare, high secretors

The Gel Microdrop (GMD) secretion assay isolates individual, viable cells based on the amount of secreted protein. The assay provides a rapid flow cytometric alternative to limiting dilution cloning. Cells are first encapsulated in a gel matrix (GelBioGel) to form GMDs. Streptavidin is added creating a bridge to bind a biotinylated 'capture' antibody. GMDs, with cells and bound antibody, are incubated for secretion and capture of the target protein. A fluorescent-labelled detection antibody is incubated with the GMDs prior to flow cytometry and sorting. The fluorescence intensity of each GMD is quantified and high producers are isolated.

Reader Service No. 111

Kersys sterilizer

From H+P Labortechnik www.hp-lab.de

Continuous sterilization of culture media

The Kersys continuous sterilizer has been developed to perform a gentle continuous sterilization of culture media for bioreactor feeds. With this system, growth media are only exposed to extreme heat for seconds, so that thermolabile components such as vitamins are not degraded.

Reader Service No. 112

Square Media Bottles

From Nalgene www.nalgenelab.nalgenunc.com

An alternative to fragile glass bottles

Moulded of clear PETG, sterile Nalgene Square Media Bottles feature a bottle neck and closure system with no liner or gasket, minimizing contamination and guaranteeing repeatable, leakproof performance. The bottles are gamma-sterilized and comply with USP 23 Class VI, are non-cytotoxic and meet USP 23 <661> requirements.

Reader Service No. 113



Square deal: Nalgene bottles up your media.



Filter tip: use this one, say Corning Costar.

Syringe filters

From Corning Costar www.corning.com

Syringe filters with a purpose

This new range of syringe filters is designed specifically for use in tissue culture. They comply with the ISO 9002 standard, are 100% integrity tested, non-pyrogenic and non-cytotoxic. The filters come in 11 sizes for volumes between 0.5 and 100 ml, with a choice of four diameters, 4, 15, 26 and 50 mm. A 50-mm, 0.22 µm Teflon vent filter is also available to protect vacuum lines and pumps or to act as an in-line filter for gases supplying CO₂ incubators.

Reader Service No. 114

Gelatin Cellware

From Biocoat www.bd.com/labware

Cell culture substrates

Biocoat Gelatin Cellware is now available from Becton Dickinson Labware. This ready-to-use format is claimed to provide reliable performance and lot-to-lot consistency. Biocoat Gelatin Cellware can be used as a cell attachment and growth promoting substrate for the culture of vascular endothelial, muscle, embryonic stem, and F9 teratocarcinoma cells. The products are stable for 12 months when stored at room temperature.

Reader Service No. 115

Tissue culture enclosures

From Labconco www.labconco.com

Three enclosures, one brochure

Protector Fibreglass Tissue Culture Enclosures are designed for tissue staining, culture growth, and virus harvesting. From the same stable are Protector PCR enclosures and Purifier HEPA-filtered PCR enclosures. All three of these recently introduced

enclosures are in an 8-page brochure, which can be ordered from the web site.

Reader Service No. 116

CO₂ incubators

From Revco revco-sci.com

Updated brochure for this incubator range

A new brochure from Revco covers the company's complete line of CO₂ incubators, including Ultima and Elite series waterjacketed incubators and the Midi size radiant wall incubator. The catalogue showcases a complete line of CO₂ incubators and accessories developed for research, clinical and life science applications. Selection includes exclusive seamless, deep drawn interior, IntrLogic microprocessor control, single (stackable) or dual chamber cabinets infrared or thermal control, HEPA filtration and personal size Midi incubators. Also included is a complete listing of product specifications and a full line of accessories. Revco designs, manufactures and markets a complete line of microprocessor-controlled, high-performance laboratory refrigerators and freezers, chest and upright low-temperature freezers from -50°C to -86°C; mechanically refrigerated -140°C and -150°C cryogenic storage systems; and automatic CO₂ cell culture incubators used in life science, industrial and hospital/clinical applications world-wide.

Reader Service No. 117

Culture tubes

From Bibby Sterilin www.bibby-sterilin.co.uk

Disposable culture tubes

Bibby's sterile culture tubes are suitable for most routine laboratory procedures and avoid the risk of cross contamination. They are biologically inert and free from mould release agents, and can be supplied with or without caps. The ribbed polypropylene caps have two positions: loose for aerobic work or sealed for anaerobic cultures. The tubes are manufactured from virgin thermoplastics — either translucent polypropylene or transparent polystyrene. The polystyrene tubes will withstand moderate centrifugation speeds and temperatures to



Disposable, but good for 121 °C.

70°C, and polypropylene tubes can be centrifuged at higher speeds and will resist temperatures from -196°C to 121°C.
Reader Service No. 118

IG650/750

From Jouan Inc. www.jouaninc.com

A new brochure describes Jouan's IG 650/750 CO₂ incubators

The IG 650 IR incubator is designed to provide rapid recovery from a door opening and humidity control for the advanced cell culture researcher. The IG 750 IR incubator adds oxygen control to the IG 650 system. The IG 650 is designed to allow culture on semisolid media and in very small volumes where evaporative losses from the media must be minimized. The IG 750 will allow the culture of cells needing either reduced oxygen levels or elevated oxygen levels.

Reader Service No. 119



CO₂ incubators: Jouan claim rapid recovery.

ECIS

From Applied BioPhysics www.biophysics.com

ECIS, electric cell-substrate impedance sensing

The ECIS biosensor measures the behaviour of cultured cells in real time and reports changes in morphology and cell movement with unparalleled sensitivity. Cells are grown in special disposable wells having small gold electrodes. With the ECIS Model 100, up to 16 individual cultures can be followed simultaneously using a weak AC signal that has no effect on the cells. A PC automatically processes all data and also directs and manages data acquisition with user-friendly software. Applications where ECIS should be considered include measurements of cell spreading and motility, determination of endothelial cell barrier function, and screening assays for *in vitro* toxicology and drug discovery.

Reader Service No. 120

Direct heat CO₂ incubators

From Forma Scientific www.forma.com

A direct approach to incubators

The latest CO₂ incubators from Forma Scientific feature heating elements and fibreglass insulation surrounding all sides of the outer chamber wall to provide uniform heating and faster recovery times after door openings. These Direct Heat CO₂ Incubators are available with either stainless steel or solid copper interior giving 6.5 cubic feet of usable culturing space and 100% coved corners for easy cleaning. A heated outer door with non-CFC insulation reduces the chance of condensation. An optional HEPA Filtered Airflow System that provides 100% HEPA filtered air throughout the chamber is available.

Reader Service No. 121



ECIS, senses cells in mid-culture.

TransferMan NK

From Eppendorf www.eppendorf.com

Cell manipulation

This new electronic micromanipulator is designed for demanding manipulation techniques involving suspension cells. TransferMan NK combines the features of a mechanical system (intuitive control of movements via proportional kinetics) with the advantages of electronic control, including storage of positions and selectable joystick transmission. TransferMan NK is particularly suitable for applications in the field of assisted reproduction (such as intracytoplasmic sperm injection (ICSI) as well as in the research lab for the transfer of embryonic stem cells (ES cells), for microinjection into suspension cells or for the latest cloning techniques. TransferMan NK can also be used for the microdissection of single cells or cell groups from fixed preparations, thus expanding the applicational range to cover the fields of pathology and cytology.

Reader Service No. 122

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

THIOREDOXIN
 GLUTAREDOXIN
 (*E.coli*, mammalian)

WWW.imcocorp.se

IMCO Corporation Ltd AB

Fax: +46 8 728 47 76 - email: info@imcocorp.se

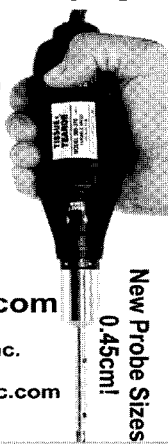
Homogenizers?

- Tissue-Tearor™
- BeadBeater™
- Mini-BeadBeater™
- BioHomogenizer™

Check
 Us
 Out

www.biospec.com

BioSpec Products, Inc.
 Ph/Fax 918-336-3363
 e-mail info@biospec.com



**Detect residual DNA with Wako's
 DNA Extractor Kit
 or Isolate Genomic DNA with
 DNA Extractor WB Kit**
 Both use the Nal Method
 No Phenol No Chloroform

Wako

Wako BioProducts
 Wako Chemicals USA, Inc.
 1600 Bellwood Road
 Richmond, VA 23237
 Telephone: (804) 271-7677
 Facsimile: (804) 271-7791
 (800) 992-9256
bioproducts@wakousa.com